Formation Pathways and Dynamics of Supermassive Black Holes in Hierarchical Galaxy Formation

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Davide Fiacconi

aus

Italien

Promotionskomitee

Prof. Dr. Lucio Mayer (Vorsitz und Leitung der Dissertation)
Prof. Dr. Romain Teyssier
Prof. Dr. Ben Moore
Prof. Dr. Philippe Jetzer

Zürich, 2016

Alla mia famiglia e ad Anna Chiara

Ama e ridi se Amor risponde
piangi forte se non ti sente
dai diamanti non nasce niente
dal letame nascono i fior

(*Via del Campo*, Fabrizio de Andrè)

Contents

List of Figures

Zusammenfassung

Basierend auf Forschungsergebnissen, welche besagen, dass schwarze Löcher fuer die Emissionen von Quasaren verantwortlich sein können, sind supermassive schwarze Löcher mittlerweile als fester und grundlegender Bestandteil massereicher Galaxien anerkannt. Doch viele Fragen ob der gegenseiten Wechselwirkung von schwarzen Löchern und Galaxien bleiben selbst nach jahrzehntelanger Forschung offen. Diese Doktorarbeit befasst sich mit einigen spezifischen Aspekten der Entstehung und dynamischen Entwicklung schwarzer Löcher im Kontext der hierarchischen Galaxienentstehung. Wir verwunden kosmologischen und idealisierten hydrodynamische Simulationen sowie analytische Techniken, um folgenden Punkte zu untersuchen:

- die Entstehung und morphologische Entwicklung der Galaxien bei hoher Rotverschiebung
- die Entstehung der Keime massiver schwarzer Löcher im Szenario des direkten Kollapses
- die Entwicklung von Paaren supermassive schwarzer Löcher und gasreichen Umgebungen

Im Detail:

Kapitel 2: Wir untersuchen die morphologische Entwicklung von 22 Galaxien bis zu $z = 3$ (Sternmassen $\sim 10^{8-11}$ $\mathrm{M}_\odot$) in der kosmologischen Zoom-In Simulation "Argo". Diese Simulation zeichnet die Entstehung eines Halo mit den Ausmassen einer Galaxiengruppe ($\sim 2 \times 10^{13}$ $\mathrm{M}_\odot$ bei $z = 0$) in einer leicht überdichten Region ab. Dies entspricht einer gewöhnlichen und daher repräsentativen Umgebung. Die sehr hohe Auflösung erlaubt uns, die Struktur der Galaxien durch mehrere Indikatoren (wie zB der Zerlegung der Dichte- und Geschwindigkeitsprofile, der B/T Verhältnisse, und der Drehimpulse) quantitativ zu untersuchen. In unserer Analyse erlangen wir vielfältige Morphologien, einschließlich "Late-Type" und unregelmäßige Scheibengalaxien mit flachen Rotationskurven, Sphäroid-dominierte "Early-Type" Scheibengalaxien, sowie eine massive elliptische Galaxie, welche bereits bei $z \sim 3$ entsteht. Weiters identifizieren wir große Zusammenschlüsse ("Major Mergers") von Galaxien als die Hauptauslöser für die Bildung von Bulgen. Kleinere Zusammenschlüsse ("Minor Mergers") und Balkenstrukturen der Sterne können in einzelnen Fällen ebenfalls das Wachstum der Bulgen vorantreiben. Wir schliessen daraus, dass morphologische Transformationen von Galaxien bei hohen Rotverschiebung

und im mittleren Massenbereiche wahrscheinlich durch ähnlich Prozesse wie bei Galaxien niedriger Rotverschiebung ausgelöst werden. Folglich kommt es zu einer frühen Entstehung der Hubble-Sequenz.

Veröffentlicht in Fiacconi, Feldmann, & Mayer 2015, MNRAS, 446, 1957.

Kapitel 3: In einigen Versionen der Theorie des direkten Kollapses kommen sogenannte Quasi-Sterne (massive Gashüllen, welche durch die Akkretion eines zentralen kleinen schwarzen Loches aufrechterhalten werden) als die letzte Phase in der Entstehung extrem massiver Sterne vor. In diesem Szenario wächst ein schwarzes Loch zunächst durch Super-Eddington Akkretion, was innerhalb von weniger als eine Million Jahren zu einem schwarzen Loch mit einer Masse von $10^4 - 10^5$ $\mathrm{M}_\odot$ führt. Solche Super-Eddington Akkretion kann jedoch auch zu starken Masseverlust führen, welcher die schlussendliche Masse des schwarzen Loches stark einschränkt. Wir betrachten die Eigenschaften von Super-Eddington Akkretion im Kontext von strahlungs- und kontinuumsgetriebenen Winden, welche von der Oberflächen massiver Objekte stammen, und wenden diese Erkenntnisse auf Quasi-Sterne an. Wir beschreiben, dass (a) photonenermüdende Effekte nicht von Relevanz sind, da der Wind von Gas-Strahlungs Enthalphie beschleunigt wird, (b) Massenabflüsse durch Winde die Masse der schwarzen Löcher auf $\sim 10^4$ $\mathrm{M}_\odot$ beschränken, und dass (c) massive Quasi-Sterne tendenziell hell und blaufarbig sind, welche sie zu potentiellen Beobachtungszielen des James Webb Space Telescopes macht.

Weiters wenden wir das Modell einer stabilen Rotation ("Steady-State Rotation") auf Quasi-Sterne an. Dabei untersuchen wir insbesondere die Verteilung des Drehimpulses in nahe der akkretierenden Region und stellen fest, dass um Quasi-Sterne mit Massen von $\gtrsim 10^5$ $\mathrm{M}_\odot$ keine zentralen Akkretionsscheiben entstehen können. Um weniger massive Quasi-Sterne jedoch können Akkretionscheiben durchaus entstehen, wobei Massenabflüsse hier jedoch so stark wären, dass das Wachstum des schwarzen Loches schnell und stark begrenzt wird. In diesem Fall gibt es zwei Möglichkeiten für das schlussendliche Schicksal der Quasi-Sternen: (a) Wenn der anfängliche supermassive Stern um die $\gtrsim 10^5$ $\mathrm{M}_\odot$ schwer ist, kann kein Quasi-Stern entstehen und der Grossteil der Masse fällt und bildet den Keim eines schwarzen Loches mit Masse $\sim 10^{4-5}$ $\mathrm{M}_\odot$, oder (b) im gegenteiligen Fall entsteht zwar ein Quasi-Sternen, jedoch wird wird die Gashülle bedingt durch Rückwirkung mit dem zentralen schwarzen Loch rasch weggeblasen, und ein kleineres schwarzen Loch mit Masse $\sim 10^{2-3}$ $\mathrm{M}_\odot$ entsteht.

Veröffentlicht in Fiacconi & Rossi 2016, MNRAS, 455, 2, und Fiacconi & Rossi, MNRAS vorgelegt.

Kapitel 4: Paare massive schwarzer Löcher mit Abständen unterhalb eines kpc sind etablierte Vorhersagen in Modellen hierarchischer Strukturentstehung. Sobald deren Abstand (zB durch Interaktion mit der einbettenden Gasscheibe) ausreichend klein wird,

verschmelzen sie und strahlen dabei starke Gravitationswellen ab. Wir analysieren, wie Umgebung klumpigen und inhomogen verteilten Gas die Dynamik der schwarzen Löcher beeinflussen kann. Zu diesem Zwecke simulieren wir zunächst Paare schwarzer Löcher mit Abständen von ~ 100 pc, welche in isolierte zirkumnukleare Scheiben eingebettet sind. Wir stellen fest, dass Interaktion durch Gravitationskräften zwischen den schwarzen Löchern und Scheiben und deren Spiralarmen, den Zerfall der Umlaufbahnen unberechenbar verändert, so dass die schwarzen Löcher in einigen Fällen langsamer als auch schneller verschmelzen. In seltenen Fällen können schwarze Löcher gar aus der Ebene der Scheibe hinausgeschleudert werden. In diesem Fall umläuft das schwarze Loch das System im meisten Fall durch den weniger dichten Sternenhintergrund, was zu längeren typischen Zeitskalen führt. Solch stochastische Effekte treten vor allem dann auf, wenn Gasklumpen existieren (mit typischen Zerfallszeiten von ~ 1 bis $\lesssim 100$ Myr), welche massiver sind als die schwarzen Löcher. Schlussendlich verifizieren wir diese Resultate mit Simulationen von verschmelzenden schwarzen Löchern mit numerischen Auflösungen von wenigen pc, wobei wir das aus Supernova-Explosionen, Sternentstehungen, und Kühlung durch Abstrahlung entstandene komplexe Interstellar Medium betrachten.

Veröffentlicht in Fiacconi et al. 2013, ApJ, 777, L14, und Roškar, Fiacconi, et al. 2015, MNRAS, 449, 494.

Abstract

Supermassive black holes are fundamental constituents of massive galaxies. This notion has been accepted since the late 1960s, when it was realised that supermassive black holes might power quasar emissions. Yet, several outstanding questions about the mutual relations between supermassive black holes and galaxies remain after several decades of research. This Thesis addresses some specific aspects of the formation and dynamical evolution of supermassive black holes in the broader context of hierarchical galaxy formation. We use cosmological and idealised hydrodynamical simulations, as well as analytical techniques, to probe:

- the formation and morphological evolution of high redshift galaxies;
- the formation of massive black hole seeds within the direct collapse scenario;
- the evolution of supermassive black hole pairs in gas-rich environments.

In detail:

Chapter 2: We study the morphological evolution of 22 galaxies down to $z = 3$ (with stellar masses $\sim 10^{8-11}$ $\mathrm{M}_{\odot}$) in the Argo cosmological zoom-in simulation. Argo traces the assembly of a group-size halo with $\approx 2 \times 10^{13}$ $\mathrm{M}_{\odot}$ at $z = 0$ in a slightly over-dense region, i.e. a common and representative environment. The high resolution allows us to quantitatively study the structure of the galaxies by means of several indicators, such as profile decomposition, B/T ratios, and angular momentum content. We recover a diversity of morphologies, including late-type/irregular disc galaxies with flat rotation curves, spheroid dominated early-type discs, and a massive elliptical galaxy, already established at $z \approx 3$. We identify major mergers as the main trigger for the formation of bulges. Minor mergers and stellar bars can also drive the bulge growth in some cases. We conclude that morphological transformations of high-redshift galaxies of intermediate mass are likely triggered by processes similar to those at low redshift and result in an early build-up of the Hubble sequence.

Published in Fiacconi, Feldmann, & Mayer 2015, MNRAS, 446, 1957.

Chapter 3: Some flavours of the direct collapse scenario postulate the existence of *quasi-stars* – massive gaseous envelopes sustained by the accretion power of a central embryo

black hole – as the final stage of supermassive stars. In the quasi-star scenario, the embryo black hole experiences an initial super-Eddington growth, that in less than a million years may leaves a 10^{4-5} $\mathrm{M}_{\odot}$ black hole seed. Super-Eddington accretion, however, may also induce vigorous mass-loss that can limit the amount of mass able to reach the black hole. We study the properties of super-Eddington, radiation continuum-driven winds launched from the surface of massive objects and we apply them to quasi-stars. We find that: (i) photon-tiring effects are not present, since the wind is ultimately accelerated by the gas-radiation enthalpy; (ii) mass outflows can severely limit the black hole growth to $\sim 10^4$ $\mathrm{M}_{\odot}$ by blowing the envelope away; and (iii) massive quasi-stars are bright and blue, therefore possible targets for the James Webb Space Telescope. We also apply a model of steady-state rotation to quasi-stars. We study the amount of angular momentum close to the accretion region. We find that quasi-stars with mass $\gtrsim 10^5$ $\mathrm{M}_{\odot}$ would not be able to form a central accretion disc. On the other hand, smaller quasi-stars can form an accretion disc, but mass outflows would be so strong to quickly halt the black hole growth. We foresee a bimodal outcome from quasi-stars: (i) when the original supermassive star is $\gtrsim 10^5$ $\mathrm{M}_{\odot}$, the quasi-star cannot form and most of the mass collapses into a black hole seed $\sim 10^{4-5}$ $\mathrm{M}_{\odot}$; (ii) in the opposite case, the quasi-star can form, but the feedback from the black hole quickly evaporates the envelope, leaving a $\sim 10^{2-3}$ $\mathrm{M}_{\odot}$ seed.

Published in Fiacconi & Rossi 2016, MNRAS, 455, 2, and Fiacconi & Rossi, submitted to MNRAS.

Chapter 4: Massive black hole pairs below kpc separations are a natural prediction of hierarchical structure formation. When their separation shrinks enough, e.g. by interacting with the surrounding gas, they coalesce with a burst of gravitational waves. We study how a clumpy and inhomogeneous gas environment may influence the pair dynamics. First, we simulate pairs at initial separation of ~ 100 pc embedded in isolated circumnuclear discs. We find that gravitational interactions with massive clumps and scattering off spiral arms erratically perturbs the orbital decay, either accelerating or decelerating it. Gravitational slingshots can occasionally kick a black hole out of the disc plane. Then, the black hole mostly orbits in the less dense stellar background, resulting in a longer orbital timescale. The stochasticity mainly emerges when there are clumps more massive than the black hole, with decay timescales ranging from ~ 1 up to $\lesssim 100$ Myr. Then, we confirm these results with a binary merger simulation at pc resolution, where a multi-phase interstellar medium develops from star formation, supernova feedback, and radiative cooling.

Published in Fiacconi et al. 2013, ApJ, 777, L14, and Roškar, Fiacconi, et al. 2015, MNRAS, 449, 494.

1

Introduction: galaxies and supermassive black holes

Abstract In this Chapter I will review the basic concepts related to cosmology, galaxy formation, and supermassive black holes. I will start by discussing the dynamical evolution of the Universe as a whole, progressively going into the details of the growth of structures first, and then setting the stage for the main ideas behind the theory of galaxy formation in a hierarchical universe. In the second part of this Chapter, I will introduce the concept of black hole in Astrophysics and the properties of supermassive black holes in the Universe, as we observe them through the radiation emitted by matter accreting onto them. I will also briefly summarise the main proposed scenarios for the formation of supermassive black holes and their evolution as black hole binaries in a galaxy merger remnants. I will conclude describing the motivations for this Thesis and giving a brief summary of its contents.

1.1 The growth of structures in a hierarchical Universe

1.1.1 Dynamical evolution of the Universe as a whole

Our deepest understanding of gravity is encoded in Einstein's theory of General Relativity. Gravity is described as the modification of the geometry of space-time under the effect of matter-energy density, whose dynamics is in turn influenced by the geometry of the space-time itself (for a review on General Relativity and its applications, see e.g. Misner, Thorne & Wheeler 1973). The essence of the concept is encapsulated in Einstein's field equations[1], which, in tensorial notation, read:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \tag{1.1}$$

where $g_{\mu\nu}$ is the metric tensor, $R_{\mu\nu}$ and R are the Ricci tensor and scalar, respectively, $T_{\mu\nu}$ is matter stress-energy tensor, and we explicitly show the contribution of the cosmological constant Λ. The other symbols assume their usual meaning (e.g. c is the speed of light in vacuum). This is a compact form for ten non-linear second-order differential equations for the unknown $g_{\mu\nu}$, which describes the geometrical properties of space-time, and ultimately gravity, as affected by $T_{\mu\nu}$.

These equations can be applied to describe the dynamics of the Universe as a whole after being further simplified according to the *Cosmological Principle: the Universe is homogeneous and isotropic on large enough scales.* As we will show below, this leads to the major prediction that the Universe expands, as also confirmed since the pioneering observations led by Edwin Hubble (Hubble, 1929). The symmetries implied by the Cosmological Principle, which is fairly true on scales larger than hundreds of Mpc, naturally select a general form for $g_{\mu\nu}$, namely the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric[2]:

$$g_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}x^{\nu} = c^2\mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1-Kr} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right], \tag{1.2}$$

where $\mathrm{d}x^{\mu} = (\mathrm{d}t, \mathrm{d}r, \mathrm{d}\theta, \mathrm{d}\phi)$ are time + spherical *comoving* coordinates, K is the curvature of the Universe, and $a(t)$ is the scale factor, which encapsulates the time evolution. K determines the overall geometrical properties of the Universe. By using a 2D equivalence between geometrical properties, (i) $K > 0$ means that the space-time corresponds to the surface of a sphere, (ii) $K = 0$ represents a flat space-time, like a 2D plane, and (iii) $K < 0$ corresponds to a 2D saddle surface.

When the FLRW metric is used, and a perfect-fluid-like stress-energy tensor $T_{\mu\nu}$ is

---

[1] These equations can be thought as Einstein's reinterpretation of the Poisson equation, $\nabla^2\phi = 4\pi G\rho$, which relates the gravitational potential ϕ to its source ρ, i.e. the matter density.

[2] We adopt Einstein's notation, i.e. summation over repeated indices is implied.

assumed[3], equation (1.1) reduces to the two *Friedmann equations* for a:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\left(\rho + \frac{p}{c^2}\right) + \frac{\Lambda c^2}{3} - \frac{Kc^2}{a^2}, \tag{1.3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{1.4}$$

where ρ and p are the rest-mass density and pressure of the matter fluid, respectively. These two equations describe the global dynamics of the Universe through a: $\dot{a} > 0$ ($\dot{a} < 0$) implies an expanding (contracting) universe, and the expansion or contraction can be either accelerated or decelerated according to the sign of $\ddot{a}$. Those features are ultimately determined by the matter-energy content of the Universe. Assuming any component with a generic equation of state of the form $p = w\rho c^2$ and plugging equation (1.4) into equation (1.3), we get:

$$\rho(a) = \rho_0 a^{-3(1+w)}, \tag{1.5}$$

with the condition $\rho(a = 1) = \rho_0$, which describes the time evolution of a general fluid whose properties are determined by w.

We can further simplify equation (1.3) by defining the critical density $\rho_{\mathrm{c},0}$, which is the total density required at $a = 1$ (i.e. at present) to have a flat Universe (i.e. $K = 0$), namely $\rho_{\mathrm{c},0} = 3H_0^2/(8\pi G)$, where $H_0 = \dot{a}/a|_{a=1}$ is the present-day expansion rate of the Universe, also called *Hubble constant*. Specifically, by normalising equation (1.3) by $\rho_{\mathrm{c},0}$, we obtain the time evolution of the Hubble parameter, i.e. the instantaneous expansion rate of the Universe $H(a)$:

$$H^2(a) = H_0^2\left(\sum_i \Omega_{i,0}\, a^{-3(1+w_i)} + \frac{\Omega_K}{a^2}\right), \tag{1.6}$$

where we define the *density parameter* of the i-th constituent of the Universe as:

$$\Omega_i(a) = \frac{\rho_i(a)}{\rho_{\mathrm{c}}(a)} = \frac{\rho_{i,0}a^{-3(1+w_i)}}{\rho_{\mathrm{c},0}H^2(a)/H_0^2} \quad \Rightarrow \quad \Omega_{i,0} = \frac{\rho_{i,0}}{\rho_{\mathrm{c},0}}, \tag{1.7}$$

and $\Omega_{\mathrm{K}} = 1 - \sum_i \Omega_{i,0}$, with the correspondences (i) $\Omega_{\mathrm{K}} > 0 \Leftrightarrow K < 0$, (ii) $\Omega_{\mathrm{K}} = 0 \Leftrightarrow K = 0$, and (iii) $\Omega_{\mathrm{K}} < 0 \Leftrightarrow K > 0$.

Equation (1.6) connects the content of the Universe with its geometry and evolution. From several observational constraints (e.g. Planck Collaboration et al., 2015), we believe that the Universe is made of: (i) **radiation**, i.e. a fluid of relativistic particles (e.g. photons) with $w = 1/3$ and $\Omega_{\mathrm{r},0} \approx 9.1 \times 10^{-5}$; (ii) **matter**, i.e. a globally pressure-less fluid with $w = 0$, made of *baryons* (i.e. ordinary matter, $\Omega_{\mathrm{b},0} \approx 0.05$) and *(cold) dark matter* ($\Omega_{\mathrm{c},0} \approx 0.26$), with a total $\Omega_{\mathrm{m},0} \approx 0.31$; and (iii) **cosmological constant**, whose

---

[3] If matter is a perfect fluid, $T_{\mu\nu} = (\rho + p/c^2)U_\mu U_\nu - pg_{\mu\nu}$, where ρ and p are the rest-mass density and pressure respectively, and U_μ is the 4-velocity field.

physical nature is still debated, that behaves as a fluid with $w = -1$ and $\Omega_{\Lambda,0} \approx 0.69$. The composition of the Universe is such that it is expanding and the expansion is currently accelerating. Moreover, the lifetime of the Universe, which is about 14.1 Gyr until today, is divided in three phases: (i) before $a_{\rm eq} = \Omega_{\rm r,0}/\Omega_{\rm m,0} \approx 2.9 \times 10^{-4}$, the Universe is dominated by radiation, which sets the scale factor $a \sim t^{1/2}$ from equation (1.6); (ii) after $a_{\rm eq}$ and until $a_\Lambda = (\Omega_{\rm m,0}/\Omega_{\Lambda,0})^{1/3} \approx 0.77$, matter dominates the energy budget, leading to $a \sim t^{2/3}$; (iii) after a_Λ the cosmological constant starts to dominate and the Universe undergoes an exponential expansion.

1.1.2 Linear and non-linear growth of perturbations in an expanding universe

Despite the utility of the Cosmological Principle, we do observe from the distribution of galaxies that the Universe is not isotropic and homogeneous on relatively small scales ($\lesssim$ 100 Mpc; e.g. SDSS survey[4]; Eisenstein et al. 2011). The inhomogeneities in the matter fluid are seeded in the very early Universe and are ultimately the places where galaxies have formed. We can describe the growth of those structures through the theory of perturbations (e.g. Coles & Lucchin 2002).

Linear perturbation theory

For the sake of simplicity, we consider Newtonian perturbations only. We generally assume that matter can be described as a self-gravitating fluid, characterised by a general barotropic[5] equation of state $p = p(\rho)$, through the equations of conservation of mass and momentum and the Poisson equation:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) &= 0, & (1.8)\\ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} &= -\frac{\nabla p}{\rho} - \nabla \Phi, & (1.9)\\ \nabla^2 \Phi &= 4\pi G \rho_{\rm tot}, & (1.10) \end{aligned}$$

where ρ, p, and $\mathbf{v}$ are the density, pressure, and velocity of the matter fluid, respectively, and Φ is the gravitational potential. $\rho_{\rm tot}$ is the total energy-matter density content of the Universe, namely $\rho_{\rm tot} = \tilde{\rho} + \tilde{\rho}_i + \ldots$, with $\tilde{\rho} = \rho + 3p/c^2$ and $\tilde{\rho}_i$'s are any other additional components (e.g. radiation). The equations above are in physical units, namely $\mathbf{r} = a(t)\mathbf{x}$ and $\mathbf{u} = \dot{\mathbf{r}} = \dot{a}\mathbf{x} + a\mathbf{v}$, where $\mathbf{x}$ and $\mathbf{v} = \dot{\mathbf{x}}$ are the *comoving* coordinates and velocities; we can transform them in comoving units by applying the transformations $\partial_t \to \partial_t - H\mathbf{x} \cdot \nabla$ and $\nabla \to \nabla/a$. Moreover, we can write the *matter* density as $\rho = \bar{\rho}(t)(1 + \delta(\mathbf{x}, t))$, where

---

[4] `http://www.sdss.org`

[5] Since the equation of state is barotropic, we consider only isentropic perturbations.

$\bar{\rho} \propto a^{-3}$ is the average, homogeneous and isotropic matter background that appears in e.g. equation (1.4), and $\delta \ll 1$ is a small perturbation; we then finally obtain[6]:

$$\begin{aligned} \frac{\partial \delta}{\partial t} + \nabla \cdot \mathbf{v} &= 0, & (1.11) \\ \frac{\partial \mathbf{v}}{\partial t} + 2H\mathbf{v} &= -\frac{c_{\mathrm{s}}^2}{a^2}\nabla\delta - \frac{1}{a^2}\nabla\Psi, & (1.12) \\ \nabla^2\Psi &= 4\pi G a^2 \bar{\rho}\delta, & (1.13) \end{aligned}$$

where $c_{\mathrm{s}}^2 = \mathrm{d}p/\mathrm{d}\rho$ is the sound speed, and Ψ is the perturbation potential associated to $\bar{\rho}\delta$. Taking the ∂_t-derivative of equation (1.11) and combining the result with equation (1.12) and (1.13), we obtain:

$$\frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} = \frac{c_{\mathrm{s}}^2}{a^2}\nabla^2\delta + 4\pi G\bar{\rho}\delta. \tag{1.14}$$

By taking the spatial Fourier transform of the latter equation, which transforms $\delta \to \delta_{\mathbf{k}}$ and $\nabla \to i\mathbf{k}$, we finally get the equation that describes the growth of density perturbations on scale $\mathbf{k}$:

$$\ddot{\delta}_{\mathbf{k}} + 2H\dot{\delta}_{\mathbf{k}} = \left(4\pi G\bar{\rho} - \frac{k^2 c_{\mathrm{s}}^2}{a^2}\right)\delta, \tag{1.15}$$

where $\dot{} = \mathrm{d}/\mathrm{d}t$. This is a second order differential equation whose solutions can be either a growing or decaying mode. We will focus on the growing modes. We can study the growth of different density perturbations at different epochs and in different conditions.

Early Universe: let us consider dark matter perturbations (i.e. characterised by $c_{\mathrm{s}} = 0$) when $a \ll 1$, such that any effect of $\Omega_{\Lambda,0}$ is negligible. Then, we have $H = H_0(\Omega_{\mathrm{m},0}a^{-3} + \Omega_{\mathrm{r},0}a^{-4})^{1/2} \equiv H_0\sqrt{\Omega_{\mathrm{r},0}}a_{\mathrm{eq}}^{-2}(1+y)^{1/2}/y^2$, where $y \equiv a/a_{\mathrm{eq}}$. Substituting this definitions in equation (1.15) and transforming $\dot{} = H_0\sqrt{\Omega_{\mathrm{r},0}}a_{\mathrm{eq}}^{-2}(1+y)^{1/2}y^{-1}\mathrm{d}/\mathrm{d}y$, we obtain:

$$\delta_{\mathbf{k}}'' + \frac{2+3y}{2y(1+y)}\delta_{\mathbf{k}}' - \frac{3}{2y(1+y)}\delta_{\mathbf{k}} = 0, \tag{1.16}$$

where $' = \mathrm{d}/\mathrm{d}y$. Very interestingly, the growing solution of this equation can be written as:

$$\delta_{\mathbf{k}}(t) = D(t)\ \delta_0(\mathbf{k}) = \left(1 + \frac{3}{2}\frac{a}{a_{\mathrm{eq}}}\right)\ \delta_0(\mathbf{k}), \tag{1.17}$$

where $D(t)$ is the linear growth factor and $\delta_0(\mathbf{k})$ is the initial perturbation. This means that, before matter-radiation equivalence, dark matter perturbations cannot grow, whereas, afterward, they grow as $\delta \propto a$, as expected in a matter-dominated universe. This result takes the name of *Meszaros effect* (Meszaros, 1974) and describes the growth of dark matter perturbation during the radiation epoch that are well within the cosmological horizon (in comoving units) $R_{\mathrm{H}}(t) = c\int_0^t \mathrm{d}t'/a(t')$.

---

[6] We also assume that the comoving velocity $\mathbf{v}$ is a small perturbation compared to the Hubble flow $\mathbf{u}_H = H\mathbf{r}$, i.e. $|\mathbf{v}|/|\mathbf{u}_H| \sim \delta \ll 1$.

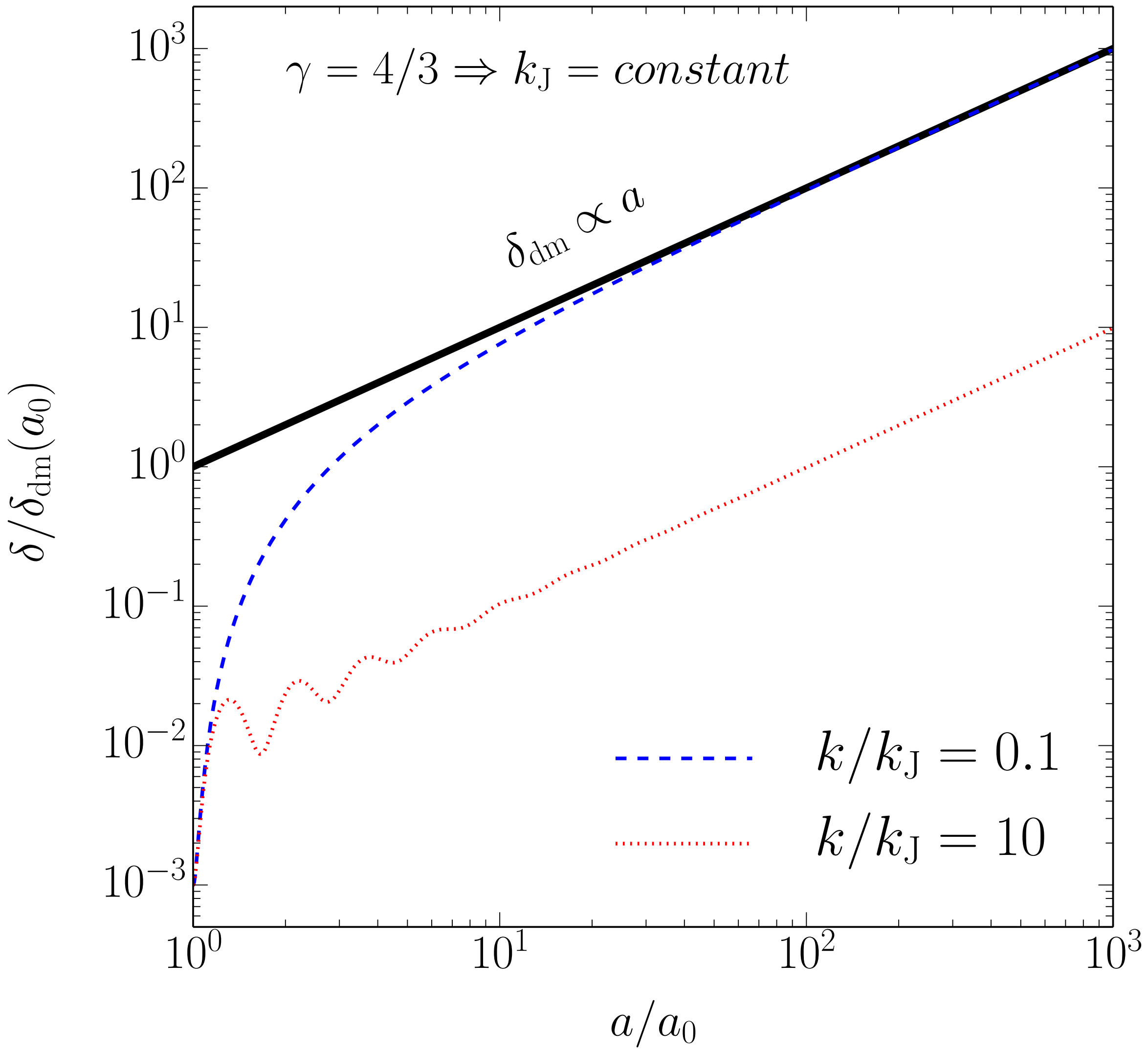

Figure 1.1 Time evolution of baryonic linear perturbations coupled to dark matter. The black thick line shows the evolution of $\delta_{\rm dm} \propto a$. The blue dashed and red dotted lines show the evolution of baryonic perturbation on scale $k/k_{\rm J} = 0.1$ and $k/k_{\rm J} = 10$, respectively. The solutions have been calculated under the assumption that $k_{\rm J}$ is constant (see the text for additional details).

Baryonic perturbations coupled to dark matter: let us focus on baryonic perturbations coupled to dark matter during the matter era. Repeating the derivation above, where the only coupling between the two components is through the gravitational potential sourced by both baryonic and dark matter, this system can be described by the following equations:

$$\ddot{\delta}_{\rm b} + 2H\dot{\delta}_{\rm b} = -\frac{c_{\rm s}^2 k^2}{a^2}\delta_{\rm b} + 4\pi G\bar{\rho}_{\rm m}\left(\frac{\Omega_{\rm b,0}}{\Omega_{\rm m,0}}\delta_{\rm b} + \frac{\Omega_{\rm m,0} - \Omega_{\rm b,0}}{\Omega_{\rm m,0}}\delta_{\rm dm}\right), \tag{1.18}$$

$$\ddot{\delta}_{\rm dm} + 2H\dot{\delta}_{\rm dm} = 4\pi G\bar{\rho}_{\rm m}\left(\frac{\Omega_{\rm b,0}}{\Omega_{\rm m,0}}\delta_{\rm b} + \frac{\Omega_{\rm m,0} - \Omega_{\rm b,0}}{\Omega_{\rm m,0}}\delta_{\rm dm}\right), \tag{1.19}$$

where $\delta_{\rm b}$ and $\delta_{\rm dm}$ are baryonic and dark mater perturbation, respectively. If we take that $\Omega_{\rm b,0} \ll \Omega_{\rm m,0}$, equation (1.19) reduces to equation (1.15) for a pressure-less fluid, whose solution we know is $\delta_{\rm dm} \propto a$ in the matter-dominated era. On the other hand, after changing the independent variable from t to a, the baryonic perturbations obey the equation:

$$\delta_{\rm b}'' + \frac{3}{2}\frac{\delta_{\rm b}'}{a} = -\frac{3}{2}\frac{k^2}{k_{\rm J}^2(a)}\frac{\delta_{\rm b}}{a^2} + \frac{3}{2}\frac{\delta_{\rm dm}}{a^2}, \tag{1.20}$$

where the Jeans scale $k_{\rm J}^2(a) = 3H_0^2\Omega_{\rm m,0}/(2c_{\rm s}^2 a)$; in general, $c_{\rm s} = c_{\rm s}(a)$. The Jeans scale describes the linear scale below (beyond) which pressure (gravity) dominates the fluid dynamics, preventing (favouring) the gravitational collapse. In the example case $p \propto \rho^\gamma$ with $\gamma = 4/3$, we have $c_{\rm s}^2 \propto a^{-1} \Rightarrow k_{\rm J} = constant$ and the solution is $\delta_{\rm b} = \delta_{\rm dm}(1 + k^2/k_{\rm J}^2)^{-1}$. Such a solution implies that baryons (during the matter-era, if the matter content is dominated by dark matter) follow dark matter perturbations, with different amplitude depending on the scale: on large scales, $k \ll k_{\rm J}$, baryons behave as a pressure-less fluid similar to dark matter, while on smaller scale, $k \gg k_{\rm J}$, the baryonic pressure is not negligible and initially tends to suppress the growth of perturbation and causes acoustic oscillations, while at later times the oscillations damp because of the expansion and perturbations grow as the dark matter, as illustrated in Figure 1.1. This behaviour is qualitatively true for any value of γ, which implies that the growth of dark matter over-densities is instrumental to guide the accumulation of the baryonic matter that would ultimately constitute galaxies.

Top-hat spherical collapse

The growth of density perturbations is guided by the interplay between gravity and cosmic expansion (and pressure, if any). At some point, they can become non-linear ($\delta \gtrsim 1$) and the following evolution is dominated by the effect of gravity.

Let us consider a homogeneous, spherical perturbation of density ρ over a background of density $\bar{\rho}$, related by $\rho = \bar{\rho}(1 + \delta)$. During an homologous collapse (i.e. before shell crossing), the mass within a shell of radius r, i.e. $M = (4\pi/3)\bar{\rho}(1+\delta)r^3$ remains constant. The evolution of such a shell can be then described in Newtonian dynamics by:

$$\frac{{\rm d}^2 r}{{\rm d}t^2} = -\frac{GM}{r^2} \quad \Rightarrow \quad \frac{1}{2}\left(\frac{{\rm d}r}{{\rm d}t}\right)^2 - \frac{GM}{r} = \mathcal{E} = \frac{H_0\Omega_{\rm m,0}}{2a}\left(\frac{{\rm d}r}{{\rm d}a}\right)^2 - \frac{GM}{r}, \tag{1.21}$$

where $\mathcal{E} < 0$ is the specific total energy of the bound over-density. In the last equality we substitute $t \to a$ assuming that we are in the matter-dominated epoch. The solution of the equation above can be expressed parametrically as:

$$r(\theta) = \frac{GM}{2|\mathcal{E}|}(1 - \cos\theta), \tag{1.22}$$

$$a(\theta) = \left(\frac{3}{2}\frac{H_0\sqrt{\Omega_{\rm m,0}}GM}{(2|\mathcal{E}|)^{3/2}}\right)(\theta - \sin\theta)^{2/3}. \tag{1.23}$$

The over-density initially expands up to a maximum radius $r_{\rm ta}$ corresponding to the turn-around time $a_{\rm ta}$ (corresponding to $\theta = \pi$) and then it collapses back at the collapse time $a_{\rm col} = 2^{2/3} a_{\rm ta}$ (corresponding to $\theta = 2\pi$). From the solution we can study the evolution of the over density, indeed we have $\bar{\rho} = \rho_{\rm c,0} \Omega_{\rm m,0} a^{-3}$ and $\rho = 3M/(4\pi r^3)$, and finally:

$$1 + \delta = \frac{\rho}{\bar{\rho}} = \frac{9}{2} \frac{(\theta - \sin\theta)^2}{(1 - \cos\theta)^3}. \tag{1.24}$$

We can compare the evolution of δ with that predicted from linear theory, i.e. $\delta_{\rm lin} \propto a$. Taylor expanding equation (1.24) for $\theta \ll 1$ (i.e. $\delta \ll 1$), we obtain $\delta_{\rm lin} \approx 3\theta^2/20$. We can use that to normalise the result of the linear theory: by Taylor expanding $r(\theta)$ and $a(\theta)$ and normalising to $a_{\rm ta}$, we obtain:

$$\delta_{\rm lin} = \frac{3}{20} (6\pi)^{2/3} \frac{a}{a_{\rm ta}}. \tag{1.25}$$

Finally, we can compare the results from the two treatments.

- **Turn around**: when $a = a_{\rm ta}$ $(\theta = \pi)$,

  $$\delta_{\rm lin} = \frac{3}{20}(6\pi)^{2/3} \approx 1.0624 \quad \Leftrightarrow \quad \delta = \frac{9\pi^2}{16} - 1 \approx 4.5517.$$

- **Collapse**: when $a_{\rm col} = 2^{2/3} a_{\rm ta}$ $(\theta = 2\pi)$,

  $$\delta_{\rm lin} = \frac{3}{20}(12\pi)^{2/3} \approx 1.6865 \quad \Leftrightarrow \quad \delta = +\infty.$$

  This result implies that an over-density whose linearly-extrapolated $\delta_{\rm lin}$ reaches 1.6865 at some time, then is likely going to collapse.

- **Virialization**: a spherical over-density does not really collapse, but it may reach the *virial equilibrium*[7], forming a bound structure called a *dark matter halo*. According to the virial equilibrium and the conservation of energy, we have $\mathcal{E} = -GM/r_{\rm ta} = \mathcal{E}_{\rm vir} = K_{\rm vir} + W_{\rm vir} = W_{\rm vir}/2 = -GM/(2r_{\rm vir})$, where $K_{\rm vir}$ and $W_{\rm vir}$ are the kinetic and potential energy after virialization, and the virial radius $r_{\rm vir} = r_{\rm ta}/2$. Therefore, at virialization, $\rho = 8\rho(a_{\rm ta})$. We can calculate from $r(\theta) = r_{\rm vir}$ when virialization occurs, and we would obtain $1 + \delta_{\rm vir} \approx 146.84$, but if we consider that virializations corresponds with the collapse at $a_{\rm col}$, than the background matter has decreased by a factor 4 compared to the turn-around value, which leads to the definition of the *virial over-density* $1 + \Delta_{\rm vir} = 18\pi^2 \approx 178$. Better approximations of the virial over-density in different cosmologies (i.e. not just matter-dominated) have been calculated by e.g. Bryan & Norman (1998).

---

[7] The virial equilibrium of a self-gravitating system is reached when $2K + W = 0$, where K and W are the kinetic and potential energy of the system, respectively.

Matter power spectrum

We have described how the over-densities evolve in both the linear and non-linear regimes. We now discuss what is distribution of amplitudes of those perturbation. By the end of inflation[8], the matter density field is expected to be seeded with gaussian perturbations on different scales, i.e. a linear superposition of normally-distributed $\delta_{\mathbf{k}}\ \forall\mathbf{k}$ (see e.g. Mo, van den Bosch & White 2010). Such a random gaussian field is fully described by its *power spectrum* $P(\mathbf{k})$, i.e. the Fourier transform of the 2-points correlation function of the matter field:

$$\langle\delta_{\mathbf{k}}\delta_{\mathbf{k}'}\rangle \propto \delta^{(3)}(\mathbf{k}-\mathbf{k}')P(\mathbf{k}), \tag{1.26}$$

where $\langle\cdot\rangle$ indicate the expectation value calculation and $\delta^{(3)}$ is the three-dimensional Dirac distribution.

After inflation, the power spectrum of the matter field is the so called primordial power spectrum (or Harrison-Zel'dovich power spectrum):

$$P_{\text{prim}}(\mathbf{k}) = A|\mathbf{k}|^n = Ak^n, \tag{1.27}$$

where A is a normalisation and the exponent n is called spectral index. Such power spectrum has the property of being scale-invariant, because the power spectrum of the gravitational potential associated to density perturbations that follow P_{prim} is independent of $\mathbf{k}$. The measured value is $n \approx 0.967$ (Planck Collaboration et al., 2015).

However, the primordial power spectrum undergoes modifications during the initial evolution of the Universe. Those modifications can be encoded in the so called *transfer function* $T(k,a)$ (e.g. Eisenstein & Hu 1999), which maps P_{prim} into the matter power spectrum P_{matter} at a reference epoch. The following evolution can be described through the linear theory (and we talk about the linear power spectrum) and is dictated by the linear growth factor $D(a)$, namely:

$$P_{\text{matter}}(k,a) = D(a)\ T^2(k,a)\ P_{\text{prim}}(k). \tag{1.28}$$

The correct normalisation of P_{matter}, which ultimately defines A, is determined observationally through the relation between the matter variance at scale R and the power spectrum:

$$\langle\delta^2\rangle \equiv \sigma^2(R,a) = \int_0^{+\infty} \frac{k^2}{2\pi^2} P_{\text{matter}}(k,a)\ \hat{W}^2(kR)\ \mathrm{d}k, \tag{1.29}$$

where $\hat{W}(Rk)$ is the Fourier transform of a filtering function (typically a top-hat filter in real space or a k-sharp filter in Fourier domain) that defines the scale R. Specifically, the observed value of $\sigma_8 \equiv \sigma(8\ h^{-1}\ \text{Mpc}, 0) = 0.816$ (Planck Collaboration et al., 2015), where

---

[8] Inflation is a brief initial phase of the evolution of the Universe (starting $\sim 10^{-36}$ s after the Big Bang and lasting for $\sim 10^{-33}$ s), when the latter experiences an accelerated exponential expansion (for details, see e.g. Coles & Lucchin 2002; Mo, van den Bosch & White 2010).

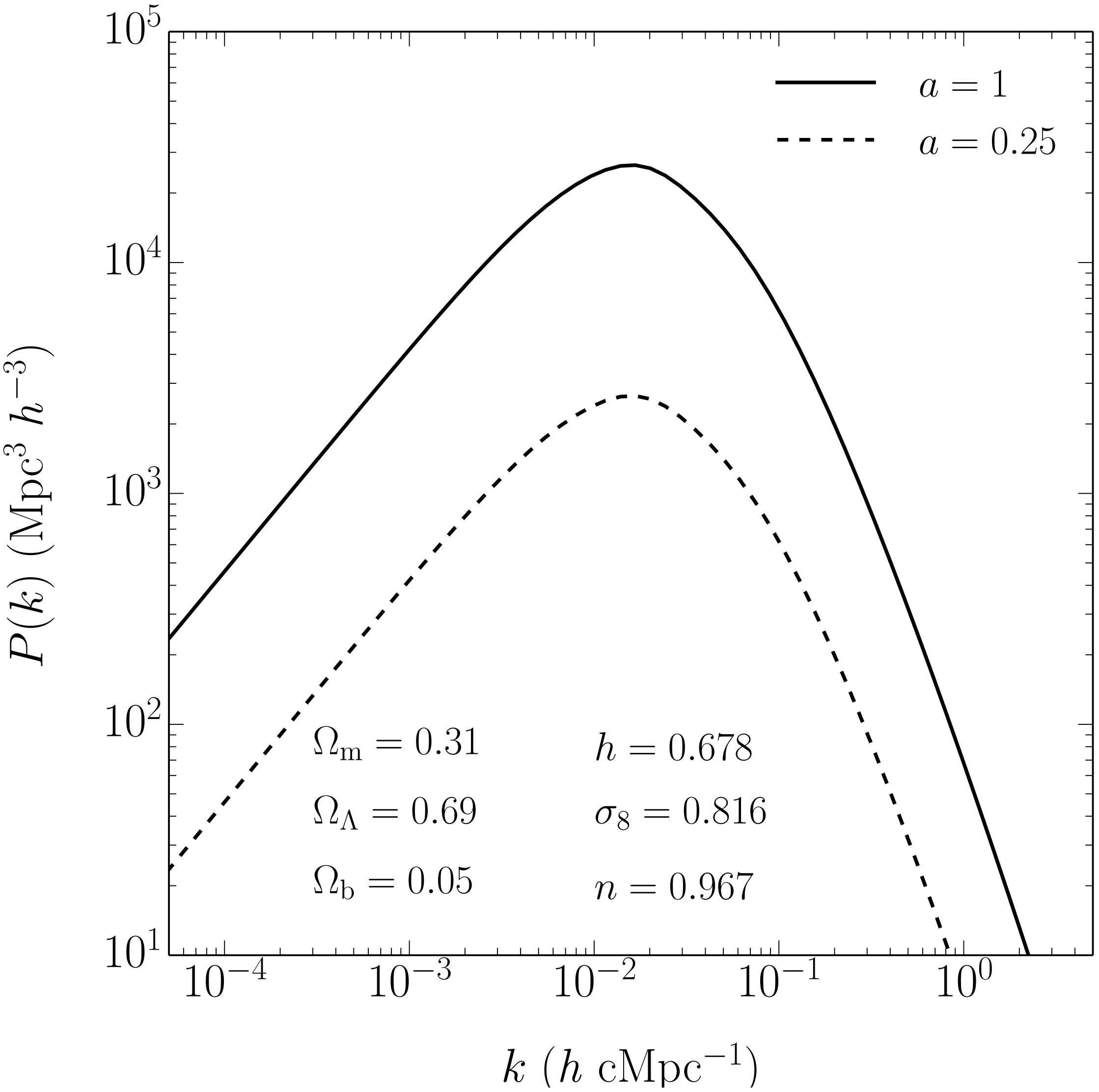

Figure 1.2 Linear matter power spectrum calculated using the transfer function from Eisenstein & Hu (1999) at $a = 1$ (solid line) and $a = 0.25$ (dashed line). We adopt a cosmology consistent with the latest results Planck Collaboration et al. (2015).

$h = H_0/(100$ km s^{-1} Mpc$^{-1})$. Figure 1.2 shows the linear matter power spectrum (calculated using the transfer function from Eisenstein & Hu 1999). Below $k \sim 10^{-2}$ cMpc^{-1}, $P_{\rm matter} \propto P_{\rm prim}$ and the evolution is still linear. Above $k \sim 10^{-2}$ cMpc^{-1}, $P_{\rm matter}$ deviates from $P_{\rm prim}$. This scale roughly corresponds to the comoving scale of perturbation entering the cosmological horizon at the radiation-matter equivalence, i.e. $R_{\rm H,eq} = 2\pi/k_{\rm H,eq} = (c/H_0)\sqrt{\Omega_{r,0}}/\Omega_{\rm m,0} \approx 41\ \Omega_{\rm m,0}^{-1}$ cMpc. Perturbations below this scale (i.e. $k \gg k_{\rm H,eq}$) undergo the Meszaros effect and they do not grow until the matter-dominated era, while super-horizon perturbations (i.e. $k \ll k_{\rm H,eq}$) continue to grow, first as $\delta \propto a^2$ before equivalence, and then as $\delta \propto a$ after equivalence. The modes that enter the horizon before equivalence have a "suppressed" growth with respect to larger scale modes; those effects are described by the transfer function and produce the large k scaling of P_{matter}.

The Zel'dovich approximation

The linear perturbation theory and the top-hat collapse are simplified models of spherical collapse of homogeneous over-densities in isolation. Knowing the matter power spectrum, a more general (i.e. non-spherical) description of the growth of (pressure-less) perturbation in the linear (and mildly nonlinear) regime can be achieved through the so called *Zel'dovich approximation* (Zel'dovich, 1970).

Since perturbations grow as $\delta(\mathbf{x},t) = D(t)\delta_0(\mathbf{x})$, we can factorise the gravitational potential through equation (1.13) as $\Psi(\mathbf{x},t) = (D/a)\Psi_0(\mathbf{x})$, where Ψ_0 satisfies the time-independent equation $\nabla^2\Psi_0 = 4\pi G\rho_{\mathrm{m},0}\delta_0$. The form of Ψ allows us to solve equation (1.12) for the velocity perturbation $\mathbf{v}$ as:

$$\mathbf{v} = -\frac{\nabla\Psi_0}{a^2}\int\frac{D}{a}\mathrm{d}t. \tag{1.30}$$

This formal solution can be better expressed after realising that, when $c_{\mathrm{s}} = 0$, D satisfies equation (1.14), which can be solved as:

$$\frac{\mathrm{d}}{\mathrm{d}t}(a^2\dot{D}) = 4\pi G\rho_{\mathrm{m},0}\frac{D}{a} \quad \Rightarrow \quad \int\frac{D}{a}\mathrm{d}t = \frac{\dot{D}}{4\pi G\bar{\rho}a}, \tag{1.31}$$

which finally leads to:

$$\dot{\mathbf{x}} = \mathbf{v} = \frac{\dot{D}}{4\pi G\rho_{\mathrm{m},0}}\nabla\Psi_0(\mathbf{x}). \tag{1.32}$$

The last equation represents the lagrangian motion of a fluid element initially at $\mathbf{x}_0$ that follows the matter flow during the collapse under the effect of gravity. We can solve the equation by linearising the right-hand side in $\mathbf{x} = \mathbf{x}_0 + \boldsymbol{\xi}$ and assuming $|\boldsymbol{\xi}| \ll |\mathbf{x}_0|$, such that we find:

$$\mathbf{x}(t,\mathbf{x}_0) = \mathbf{x}_0 + \boldsymbol{\xi}(t) = \mathbf{x}_0 + \frac{D(t)}{4\pi G\rho_{\mathrm{m},0}}\nabla\Psi_0(\mathbf{x}_0). \tag{1.33}$$

Figure 1.3 shows the result of a calculation using the Zel'dovich approximation in 2 dimensions (with cosmological parameters from Planck Collaboration et al. 2015). The results have been extended into the nonlinear regime, when shell crossing occurs and the Zel'dovich approximation is no longer valid. However, in the linear regime, this approach is very powerful to describe the matter density field and is also used to initialise numerical calculations by prescribing the initial positions of particles, meant to represent discrete lagrangian elements of the matter fluid

1.1.3 The formation of galaxies: general concepts

During the evolution of the Universe, perturbations in the dark matter density field can grow, in particular after radiation-matter equivalence. After a phase of linear growth, their

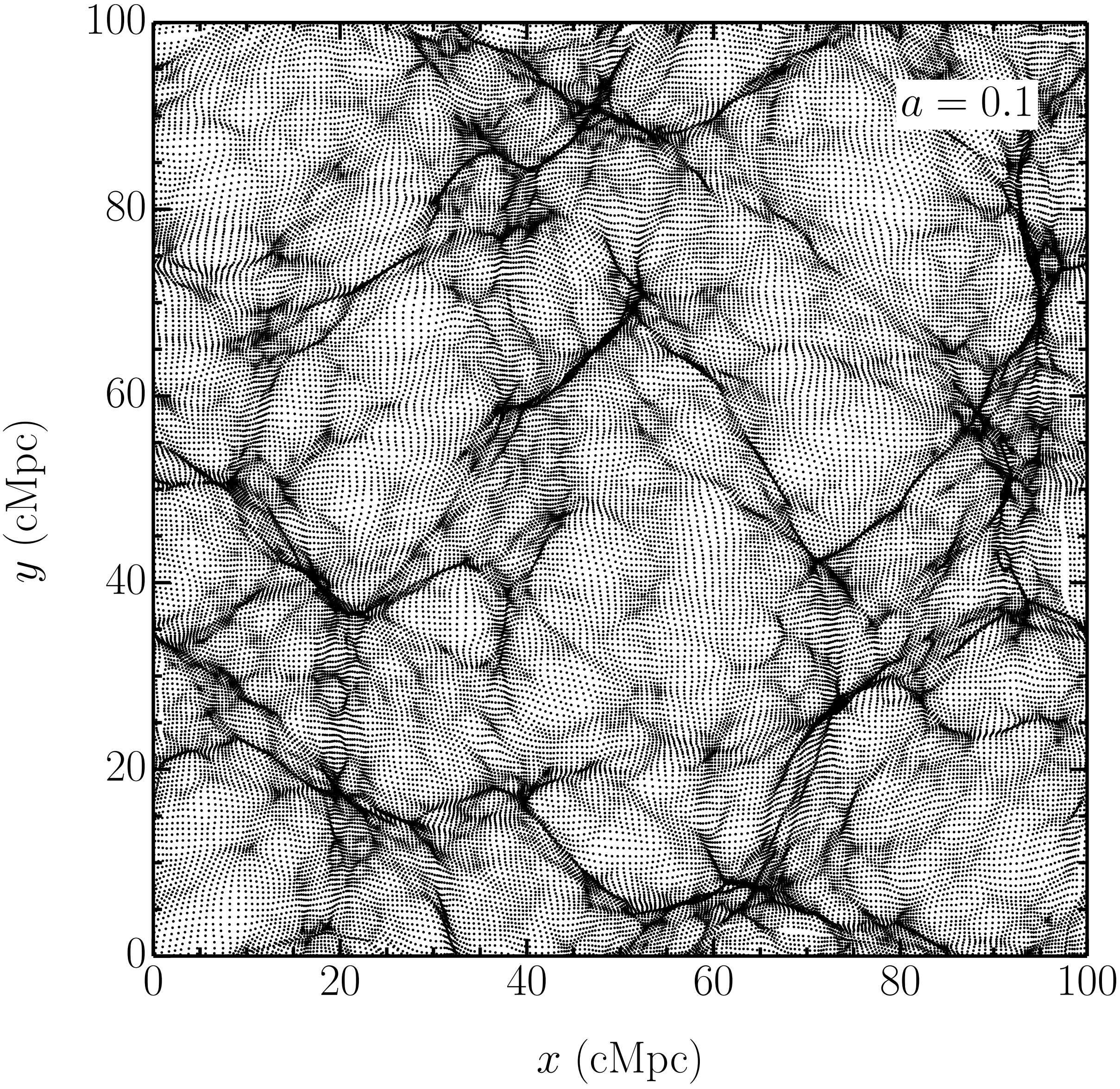

Figure 1.3 Two-dimensional Zel'dovich approximation in a Planck Collaboration et al. (2015) cosmology at $a = 0.1$, starting from $a = 0.01$. The box size is 100 comoving Mpc.

evolution becomes non-linear and mostly driven by gravity, which makes them collapse until they can reach the virial equilibrium, forming dark matter haloes (see Section 1.1.2).

Dark matter haloes are nearly spherical (but generally triaxial) bound concentrations of dark matter, extensively studied in their properties through N-body simulations able to follow all their non-linear dynamics and growth (e.g. Navarro, Frenk & White 1996, 1997; Springel et al. 2005). They follow a nearly universal spherical density profile, called Navarro-Frenk-White (NFW; Navarro, Frenk & White 1997) profile:

$$\rho(r) = \frac{\rho_0}{(r/r_\mathrm{s})(1 + r/r_\mathrm{s})^2}. \tag{1.34}$$

This is a combination of two power-laws: below a scale radius r_s, $\rho \propto r^{-1}$, whereas above

$r_{\rm s}$, it goes as $\rho \propto r^{-3}$, until it reaches the virial radius, defined as the radius that encloses an average density Δ times larger than the background (or critical) density. Δ can be either ~ 200 or $a(t)$-dependent (Bryan & Norman, 1998). The halo extends to the virial radius $R_{\rm vir}$; it contains the virial mass $M_{\rm vir} = (4/3)\pi\Delta\rho_{\rm c}R_{\rm vir}^3$, and has a virial velocity $V_{\rm vir}^2 = GM_{\rm vir}/R_{\rm vir}$.

As discussed above, baryons tend to follow the collapse of dark matter over-densities. The commonly accepted idea is that such behaviour is at the base of the formation of galaxies (e.g. White & Rees 1978; Fall & Efstathiou 1980; White & Frenk 1991). The baryons, initially in the form of gas, enter the virial radius of dark matter haloes that represent gravitational potential wells unaffected by the presence of gas. The gas gets shock-heated up to the virial temperature, i.e. the equivalent temperature corresponding to the total energy when the system is in virial equilibrium. Then, a fraction of the gas, which represents a fraction $m_{\rm d}$ of the halo mass, favoured by the radiative cooling that locally decreases the pressure support, accumulates at the centre of the halo until it condensates enough at the centre to become self-gravitating. Then, the detailed processes of star formation from gas will eventually be responsible for the formation of the central galaxy.

An important feature that we have not discussed in the simplified treatment of Section 1.1.2 is that during the collapse, dark matter haloes acquire angular momentum due to larger-scale gravitational torques. The total angular momentum J of a halo is described by the spin parameter $\lambda = J|E|^{1/2}G^{-1}M_{\rm vir}^{-5/2}$, where $|E|$ is the binding energy of the halo; it has been found that λ follows a log-normal distribution with mean $\bar{\lambda} \approx 0.05$ and dispersion $\sigma_\lambda = 0.5$ (e.g. Warren et al. 1992). Such torques also affect the gas, which in turn may retain a fraction $j_{\rm d}$ of the angular momentum of the halo during the collapsing phase toward the centre of the halo. The interplay between radiative cooling and centrifugal balance makes the gas settle into a thin-disc configuration (Fall & Efstathiou, 1980; Mo, Mao & White, 1998). If we assume that the disc has an exponential surface density profile (as often observed for the surface brightness of disc galaxies), i.e. $\Sigma(R) = \Sigma_0 \exp(-R/R_{\rm d})$, the constraints of mass conservation (disc mass $M_{\rm d} = m_{\rm d}M_{\rm vir}$) and angular momentum conservation (disc angular momentum $J_{\rm d} = j_{\rm d}J$) in a NFW halo lead to the scaling at $a = 1$ (Mo, Mao & White, 1998):

$$\begin{aligned} R_{\rm d} &= \frac{1}{\sqrt{2}}\frac{j_{\rm d}}{m_{\rm d}}\lambda R_{\rm vir}\frac{f_R}{\sqrt{f_c}} \approx \\ &\approx 5.8\left(\frac{\lambda}{0.05}\right)\left(\frac{M_{\rm vir}}{10^{12}\ {\rm M}_\odot}\right)^{1/3}\left(\frac{\Delta(a)}{94}\right)^{-1/3}\left(\frac{H(a)}{H_0}\right)^{-2/3}\ {\rm kpc}, \end{aligned} \tag{1.35}$$

where we assume $m_{\rm d} = j_{\rm d} = 0.05$, $f_c \approx 2/3 + (c/21.5)^{0.7} \approx 1.25$ for $c = 10$, and $f_R \approx (\lambda/0.1)^{-06+2.71m_{\rm d}+0.0047/\lambda}(1 - 3m_{\rm d} + 5.2m_{\rm d}^2)(1 - 0.019c + 0.00025c^2 + 0.52/c) \approx 0.68$ for the same choice of parameters. From this, Σ_0 and the circular velocity curve can be specified, capturing the gross physical processes that determines the formation of disc galaxies when compared statistically with observed samples (Mo, Mao & White, 1998).

This qualitative and simplified picture has been largely updated in the last 25 years. Although the gross idea still holds, there are substantial differences in the way gas can reach the centre of dark matter haloes of different masses. Gas effectively shocks at about the viral radius when it accretes onto haloes with mass larger than $\sim 10^{11-12}$ $\mathrm{M_\odot}$ (Birnboim & Dekel, 2003; Dekel & Birnboim, 2006). However, at smaller masses, smooth accretion is dominated by cold flows, i.e. anisotropic and filamentary accretion flows of cold ($\sim 10^4$ K) gas that penetrate never shocked the virial volume, possibly down to the central galaxy (Kereš et al., 2005; Dekel & Birnboim, 2006; Kereš et al., 2009). This has been often dubbed as the "cold mode of accretion", in contrast to the "hot mode of accretion" represented by the inflow of post-shock gas at the viral temperature that cools well within the Hubble time. Cold accretion has been found also to dominate gas accretion onto galaxies at redshifts $z > 2-3$, whereas it decreases sharply in small galaxies at lower redshifts $z \lesssim 1$ (Kereš et al., 2009). On the other hand, hot accretion contributes little over time, but it becomes important at $z \lesssim 1$ for larger haloes. Figure 1.4 shows examples of cold flows at $z = 2$ from a cosmological simulation and how they change with halo masses, being able to penetrate till the centre in lower mass ($10^{11.5}$ $\mathrm{M_\odot}$) haloes, and breaking into small cold blobs in higher mass ($10^{12.5}$ $\mathrm{M_\odot}$) haloes. The implications of the dominance of cold accretion are broad, since it would ultimately be the driving mechanism of star formation in galaxies. However, the actual formation of galaxies, and in particular of the development of their structure and morphology, does not just passively follow the rate of gas accretion, but it is also largely governed by small-scale physical processes, such as radiative cooling, star formation and stellar feedback. Specifically, stellar feedback in different forms (e.g stellar winds, stellar radiation, and energy release from supernovae) has been proved to be of fundamental importance in regulating the rate of gas-to-star conversion and the retain of fresh gas still infalling from larger scales at halo masses below $\sim 10^{12}$ $\mathrm{M_\odot}$ (e.g. Navarro & Steinmetz, 2000; Thacker & Couchman, 2001; Governato et al., 2004; Mayer, Governato & Kaufmann, 2008; Governato et al., 2010; Agertz et al., 2013, and references therein). Despite those mechanisms are broadly understood in their general features, they are still under study to assess which are the dominant ones in shaping galaxies. On the other hand, different feedback mechanisms likely come into play at larger halo masses. Although the debate is still open, the feedback from supermassive black holes (see Section 1.2.3) is a possible candidate to regulate star formation and gas acquisition in galaxies at the centre larger haloes hosting groups and clusters of galaxies.

Another fundamental aspect that we implicitly neglected in Section 1.1.2 is that haloes do not grow in isolation, but they are influenced by the environment. Indeed, the distribution of perturbations over different spatial (and mass) scales is such that smaller

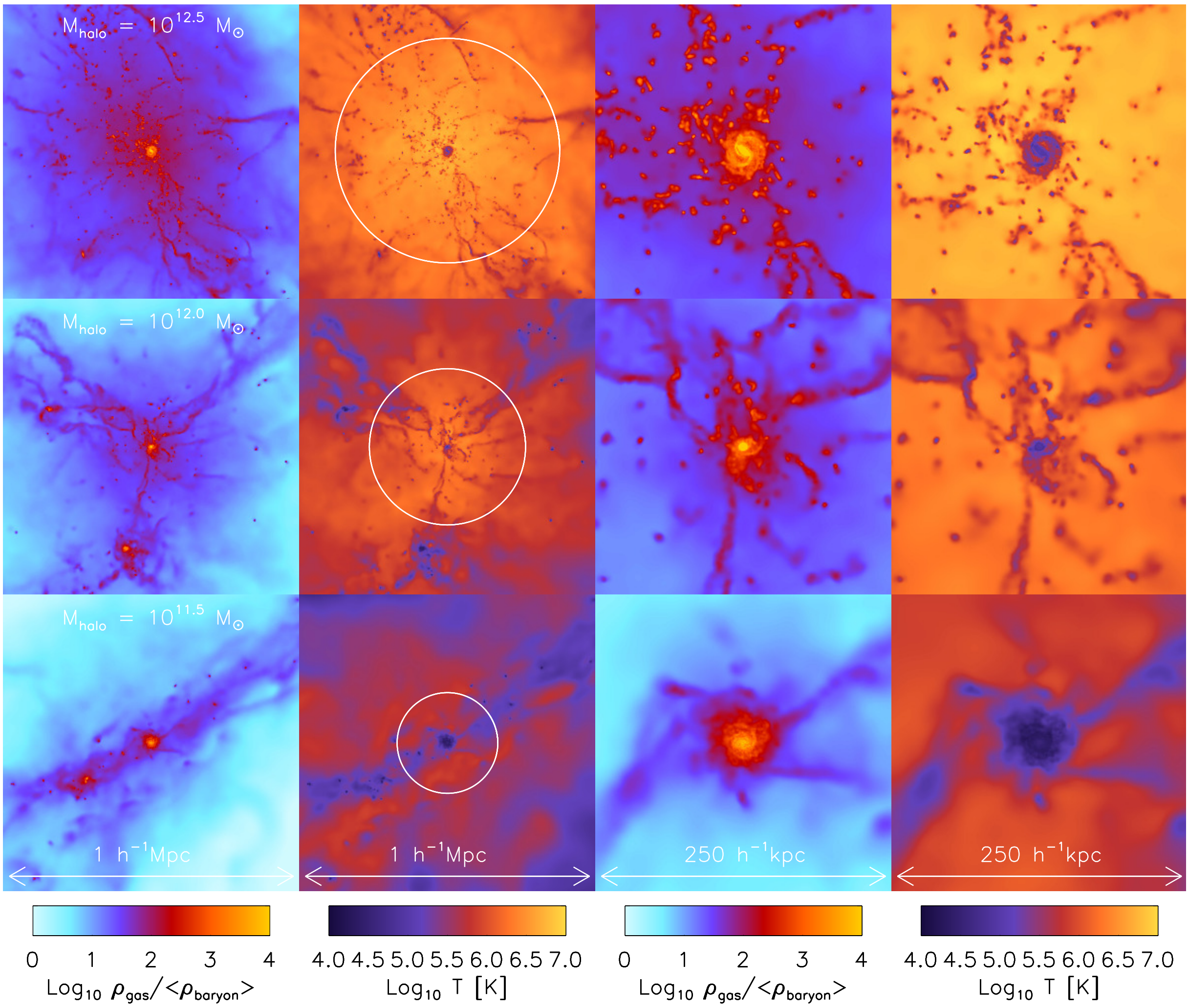

Figure 1.4 Density and temperature maps of haloes with decreasing mass from top to bottom at $z = 2$ in the Overwhelmingly Large Simulation. Cold flows are clearly visible the temperature maps as filaments entering the viral radius, marked by the white circle in the second column. Adopted from van de Voort et al. (2011).

over-densities grow first and become the building blocks of larger scale perturbations, which form and grow because of the merging of smaller haloes (White & Frenk, 1991). When merging haloes already contain galaxies, the latter also undergo a merger, which is “major” or “minor” whether the mass ratio is close to unity or $\gtrsim 4$, respectively. A single halo experiences many mergers (mostly minor) during its evolution; its merger history also influences the properties of the final galaxy since major mergers, when frequent enough, tend to destroy fragile structures like thin discs and to boost the consumption of gas into stars, leading to the formation of elliptical galaxies (e.g. Barnes 1992; Hernquist, Spergel & Heyl 1993; Barnes & Hernquist 1996; Naab, Khochfar & Burkert 2006).

Although broadly clear and commonly accepted in its general principles, the research panorama of galaxy formation and evolution is still an highly active area of research that can be hardly reduced to simple rules because of the high complexity and non-linearity of the involved processes, both on cosmological as well as sub-galactic scales. Indeed, current research strongly relies on tools such as numerical simulations and semi-analytical models to couple the many involved processes, though often treated in a simplified and phenomenological fashion. Nonetheless, the broad ideas presented here, that are at the basis of many analytical models, qualitatively describe the current general picture of galaxy formation.

1.2 Supermassive black holes in the Universe

1.2.1 Astrophysical black holes: general remarks

Another fundamental prediction of General Relativity is the existence of black holes. Black holes are space-time singularities that are able to induce a gravitational field strong enough that nothing, not even light, can escape from them. Specifically, the region from which no information can leak out is contained within the *event horizon*. According to the no-hair theorem, the only properties that can be inferred from outside the event horizon through their influence on the surrounding space-time are the *mass*, the *angular momentum* (if rotating) and the *charge* (if non-neutral) of the black hole. Any additional information (for which hair is a metaphor) on the matter/energy contained within the event horizon is not accessible from outside.

For most astrophysical purposes, black holes can be regarded as spherical distributions of mass $M_\bullet$ and radius:

$$R_{\mathrm{S}} = \frac{2GM_\bullet}{c^2} \approx 3\times 10^5 \left(\frac{M_\bullet}{\mathrm{M}_\odot}\right)\ \mathrm{cm}. \tag{1.36}$$

R_{S} is the *Schwarzschild radius*: it represents the size of the event horizon, which in fact is a spherical surface, of a non-rotating neutral black hole (Schwarzschild, 1916). While we can safely neglect any effect from the charge, black hole angular momentum may be relevant in some circumstances. This is usually parametrised through the *spin parameter* $a = |\mathbf{J}|c/(GM_\bullet^2)$, where $\mathbf{J}$ is the angular momentum of the black hole, and $0 < a < 1$ (Kerr, 1963). When $a > 0$, $\mathbf{J}$ defines a symmetry axes around which the black hole (and the space-time) rotates; the most notable effect is the so called *frame dragging*: loosely, a test particle close enough to the black hole cannot avoid to rotate as the black hole does. This occurs between the event horizon, i.e. $R_{\mathrm{hor}} = R_{\mathrm{S}}(1+\sqrt{1-a^2})/2$, and the *ergosphere*, whose section in a plane containing $\mathbf{J}$ is $R_{\mathrm{es}}(\theta) = R_{\mathrm{S}}(1+\sqrt{1-a^2\cos^2\theta})/2 \geq R_{\mathrm{hor}}$, where θ is the colatitude angle. The horizon is spherical, while the ergosphere goes from R_{hor} on the poles to R_{S} on the equator.

We can identify two classes of black holes: stellar black holes and supermassive black holes[9]. The major difference is in their mass: stellar black holes typically range from

---

[9] For the sake of completeness, we mention here that other two classes of black holes have been

$\gtrsim 3$ $M_\odot$ to ~ 20 $M_\odot$, while supermassive black holes have masses $\gtrsim 10^6$ $M_\odot$. This striking difference suggests that those objects have a different origin, and indeed they are also detected in different environments. Stellar black holes may form during the last evolutionary stage of stars $\gtrsim 25$ $M_\odot$, when the core collapses and triggers a supernova explosion (e.g. Woosley, Heger & Weaver 2002); they are spatially associated with star-forming regions. On the other hand, the formation of supermassive black holes is still debated (e.g. Volonteri 2010), but it is widely accepted that they reside in the central core of the most massive galaxies and they power the emission of Active Galactic Nuclei (AGNs) through accretion. In the following, we will solely focus on supermassive black holes, introducing the basic concepts of accretion and the role of supermassive black holes in AGNs and galaxy evolution.

Active galactic nuclei: a historical perspective

Soon after it has been accepted that galaxies were objects similar but external to the Milky Way, Seyfert (1943) first systematically studied nuclear spectra of nearby spiral galaxies with prominent nuclear emission. He found that some galaxies were hosting forbidden emission lines more common to planetary nebulae than stars. Specifically, they were divided in two kinds: (i) spectra with very broad ($\sim 10,000$ km s^{-1}) permitted lines (i.e. hydrogen lines) and bright narrow forbidden lines, and (ii) spectra with narrow permitted and forbidden lines. Those observations have not been appreciated enough until the discovery of "quasars" (or Quasi Stellar Objects; QSOs).

During the late 50's, the third Cambridge (3C) radio survey (Edge et al., 1959; Bennett, 1962) collected the positions and fluxes of many radio sources. Among them, 3C 295 appeared to have a very high measured redshift[10] $z \approx 0.46$ (Minkowski, 1960), and 3C 48 was shown to coincide with a 16 mag variable star embedded in a surrounding nebulosity, characterised by a blue spectrum with broad emission lines at unknown wavelengths. The puzzle of the strange lines was solved by analysing the spectrum of 3C 273, a radio source divided in two distinct components A and B, which have been found to coincide with a 13 mag star and a faint jet pointing away from it, respectively. Specifically, Schmidt (1963) recognised four hydrogen lines of the Balmer sequence at the redshift $z \sim 0.16$, and contemporarily Greenstein (1963) measured the redshift $z \sim 0.37$ of 3C 48 from the same lines, both concluding that the redshifts of the sources were likely cosmological. The inferred distance for 3C 273 implied that the optical luminosity were much larger than that of giant ellipticals, the nuclear emission were coming from less than 1 kpc, and the jet

---

proposed to exist: intermediate mass black holes (with masses $\sim 10^3 - 10^5$ $M_\odot$), and primordial black holes (i.e. formed because of primordial over-densities after inflation). However, none of them have been demonstrated to exist so far.

[10] The cosmological redshift of a source is the red-ward displacement of emission lines λ originally at λ_0, namely $z = \lambda/\lambda_0 - 1$. It is associated to cosmological expansion and it relates to the expansion factor as $1 + z = 1/a$.

projected length were about 50 kpc. Moreover, there were photometrical hints that most of the emission was nuclear, with a fainter, extended source (i.e. a galaxy) surrounding it, suggesting a similarity with the local galaxies observed by Seyfert (1943). Greenstein & Schmidt (1964) concluded that a mass $\gtrsim 10^9$ $M_\odot$ within 1-10 pc was required to produce enough energy to sustain the emission, surrounded by a more extended region responsible for the radio counterpart. This kind of radio stellar-like objects are called quasars since then.

Soon after the first properties of radio quasars were identified, Sandage (1965) reported the discovery of a large population of objects very similar to quasars, but without radio emission. Specifically, they were characterised by an "ultraviolet excess" when compared with normal stars in colour-colour diagrams. Such an excess is also common to quasars and allows to quickly recognise them in optical surveys. Moreover, Sandage (1965) noted that most of those sources had apparent magnitudes > 15, likely consistent with extragalactic sources (as confirmed by spectra showing redshifts as high as $z \sim 1$) and that they outnumbered the radio-loud quasars possibly by a factor of $\gtrsim 100$.

At this point, most of the observational properties of quasars and Seyfert galaxies (or, more generally, of AGNs) were recognised. The nature of the powering mechanism remained an outstanding question until Salpeter (1964) and Zel'dovich (1964) proposed that accretion onto a supermassive black hole might be responsible for AGN emission.

Accretion theory: basic concepts

We briefly review the basic ideas behind accretion the theory. Accretion is a way to extract gravitational potential energy from a massive and compact body. Let us consider a star of mass M and radius R. Suppose that a matter parcel m is free falling onto M from infinity with initial velocity $v = 0$. The conservation of total energy imposes that $v^2/2 - GM/r = 0$, where r is the distance between m and the centre of M. Therefore, at any position r, the parcel will have a kinetic energy $K = mv^2/2 = GMm/r$, which progressively increases as $r \to R$. When the parcel hits the stellar surface, it reaches the maximum kinetic energy $K_{\rm max} = GMm/R$ and then it suddenly stops, joining the stellar matter. This implies that $K_{\rm max}$ has to be dissipated, likely as heat on the stellar surface that can be then radiated away. The radiated energy $E_{\rm rad} = K_{\rm max}$ can be expressed as a fraction η of the rest-mass energy of the infalling parcel:

$$E_{\rm rad} \equiv \eta mc^2 = \frac{GM}{Rc^2}mc^2 \quad \Rightarrow \quad \eta = \frac{GM}{Rc^2} = \frac{1}{2}\frac{R_{\rm S}}{R}. \tag{1.37}$$

When we consider a compact object, such as e.g. a neutron star, the *radiative efficiency* η can be very large, namely $\eta \approx 0.22\ m_{1.5}\ r_{10}^{-1}$, where $M = 1.5\ m_{1.5}$ $M_\odot$ and $R = 10\ r_{10}$ km. This is much larger than e.g. the efficiency of the nuclear reaction $4{\rm H} \to {}^4{\rm He}$, where only 0.7 % of the rest-mass gets converted into energy.

In the previous example we considered a radial motion and an object with a solid surface. However, accreting matter would likely have some specific angular momentum L, and there will be a radius R_c where the parcel will orbit on a circle around the central mass M. This radius is given by $L = R_\mathrm{c}V_\mathrm{c}(R_\mathrm{c}) = \sqrt{GMR_\mathrm{c}}$; we assume that the parcel orbits at R_c on a nearly circular orbit and that is has an initial mechanical energy $E_\mathrm{in} = mV_\mathrm{c}^2(R_\mathrm{c})/2 - GMm/R_\mathrm{c} = -GMm/(2R_\mathrm{c})$. The parcel will just move on its circular orbit at R_c unless there is a process that removes angular momentum from the particle. We suppose that such a process occurs very slowly, regardless of its specific nature, and that it makes the parcel shift from larger to smaller *circular orbits*. This continues until $r = R$, where the final energy is $E_\mathrm{fin} = -GMm/(2R) < E_\mathrm{in}$. The dissipated energy is radiated away and it is $E_\mathrm{rad} = E_\mathrm{in} - E_\mathrm{fin} = GMm/(2R)(1 - R/R_\mathrm{c}) \approx GMm/(2R)$ if $R_\mathrm{c} \gg R$. Finally, we obtain the radiative efficiency for the case of accretion with angular momentum:

$$\eta \approx \frac{GM}{2Rc^2} = \frac{1}{4}\frac{R_\mathrm{S}}{R}. \tag{1.38}$$

The main difference between the two cases is that the latter is valid also for black holes, while the former is not. Indeed, the horizon surface, which is the ultimate surface that an accreting particle can reach, is not a physical surface and a radially accreting parcel would just flow through, without dissipating energy at the surface. Moreover, it is worth to note an implicit but fundamental assumption, namely that the initial radius R_c is much larger than the final radius R till which the material is able to radiate, because otherwise it would not be possible to extract radiative energy from the infalling matter. This is particularly relevant for black holes, in which case R corresponds to the radius of the innermost stable circular orbit $R_\mathrm{isco} \geq R_\mathrm{hor}$; within R_isco, the infalling matter unavoidably follows plunging orbits. The value of R_isco depends on the spin parameter a: $R_\mathrm{isco} = 3R_\mathrm{S}$ when $a = 0$ (i.e. for Schwarzschild black holes) and it approaches $R_\mathrm{isco} = R_\mathrm{S}/2$ as $a \to 1$ and matter is co-rotating, or it approaches $R_\mathrm{isco} = 9R_\mathrm{isco}/2$ otherwise (i.e. for counter-rotating matter or equivalently when $a \to -1$). The process responsible for the redistribution of angular momentum is thought to be viscosity produced by instabilities due to the presence of magnetic fields in the rotating accretion disc (Balbus & Hawley, 1991).

Finally we note that the same line of reasoning can be applied when a mass flow $\dot{M}$ is considered, in which case it would produce a luminosity $L = \eta\dot{M}c^2$. However, this luminosity may interact with the accreting flow while streaming through when it exceeds the so called *Eddington limit*. Consider a generic plasma of (ionised) hydrogen surrounding a central mass M: protons (i.e. the most massive component) will be gravitationally attracted by the force $F_\mathrm{g} = -GMm_\mathrm{p}/r^2$, while electrons will be scattered outward by radiation coming from the central source of luminosity L trough the radiative force $F_\mathrm{r} = \kappa_\mathrm{es}L/(4\pi r^2c^2)$, where $\kappa_\mathrm{es} \approx 0.35~\mathrm{cm}^2~\mathrm{g}^{-1}$ is the electron scattering opacity. Balancing the two forces, we find that the gas can be pushed outward when $F_{\rm r}$ outweighs $F_{\rm g}$, i.e. when:

$$L > L_{\rm Edd} \equiv \frac{4\pi G M c}{\kappa_{\rm es}} \approx 1.4 \times 10^{38} \left(\frac{M}{\rm M_{\odot}}\right) \ {\rm erg\ s^{-1}}, \tag{1.39}$$

which depends only on the mass of the central object, i.e. the source of gravity. When the accretion luminosity from the central object exceeds $L_{\rm Edd}$, the photons diffusing through the gas are in principle enough to halt the inflowing gas motion because the latter would not be dominated by gravity anymore. This in turn may stop accretion, and therefore reduces the pushing luminosity up to the point where gravity starts to dominate again and triggers new accretion.

1.2.2 Observations of supermassive black holes

The current paradigm supports the idea that different observational classes of AGNs (e.g. Seyfert galaxies, quasars, radio-galaxies, etc.) are all powered by a supermassive black hole accreting gas through an accretion disc typically smaller than $10^{-2} - 10^{-3}$ pc. This is in turn surrounded by gas clouds which divides in Broad Line Regions and Narrow Line Regions at distances ~ 0.1 pc and < 1 kpc, respectively. As their names suggest, they are responsible for the broad and narrow permitted and forbidden emission lines in AGNs' spectra. The central region is also surrounded by a dusty, obscuring torus with size ~ 1 pc. Finally, some AGNs are radio-loud, often hosting jets, whose production is likely related to the extraction of rotational energy from a spinning black hole with a magnetised accretion disc (Blandford & Znajek, 1977). This structure is schematically represented in Figure 1.5. The different AGN classes are ultimately related to the presence or not of radio-loudness/jets and to different inclinations of the line of sight through the central region (Antonucci, 1993; Urry & Padovani, 1995).

As previously mentioned, the main property of a black hole is its mass. Therefore, estimating the mass of supermassive black holes possibly powering AGNs is an indirect observational evidence for the existence of supermassive black holes and for the nature of AGNs (for a review, see Peterson 2014). Several methods have been developed, mostly divided in "direct" and "indirect". We focus here on the first kind. Direct methods derive the mass of a supermassive black hole through direct effects of the the black hole on the dynamics of the surrounding gas or stars. When the resolution is enough, it is possible to resolve the motion of individual sources around the supermassive black hole. Those are stars for the Milky Way and water maser emission from the dusty torus of NGC 4258, which shows a Keplerian motion consistent with a central point-mass of $M_{\bullet} = 4.1\times10^{6}$ M$_{\odot}$ and $M_{\bullet} = 3.8\times10^{7}$ M$_{\odot}$, respectively. Otherwise, it is possible to measure $M_{\bullet}$ by modelling the cumulative motion of stellar or gaseous systems around the black hole through the superposition of stellar/gaseous orbits in the galactic potential in order to model the emission/absorption line width and shape.

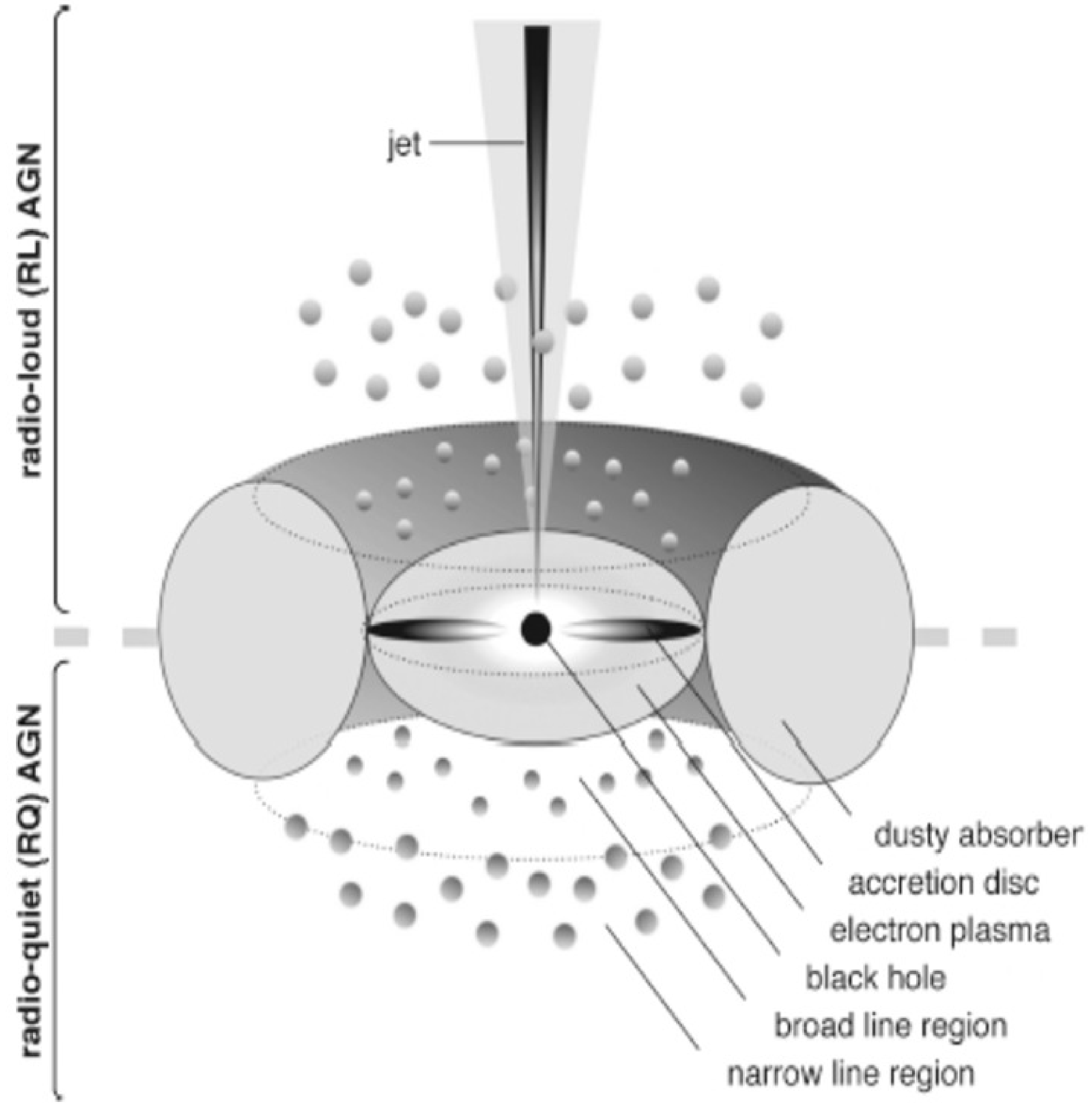

Figure 1.5 Schematic representation of the central region of an AGN (figure taken form Beckmann & Shrader 2012).

Another method is the so called "reverberation mapping" (Peterson, 1993). The main idea is that it is possible to determine the size of the Broad Line Region as $R_{\rm BLR} = c\, t_{\rm rev}$, where $t_{\rm rev}$ is the time lag between correlated fluctuations in the spectrum continuum (i.e. in the emission from accretion onto the central black hole) and in the line intensity (i.e. the response of the clouds to the time-varying illumination from the central source). Than, the velocity of the clouds at $R_{\rm BLR}$ can be determined from the width ΔV of the associated line, with geometrical corrections. Finally, the black hole mass can be inferred from the virial equilibrium as $M_\bullet \sim R_{\rm BLR} \Delta V^2/G$. This method has been successfully applied to several local galaxies, and even to slightly higher redshift galaxies through the calibration of an empirical relation between $R_{\rm BLR}$ and the continuum luminosity at 5100 Å (Bentz et al., 2006).

1.2.3 Scaling relations and AGN feedback: a sign of coevolution?

As soon as several supermassive black hole mass estimates were available, it has been immediately recognised that they correlate with some of the properties of the host galaxies, such as the bulge mass $M_{\rm b}$ (e.g. Magorrian et al. 1998; Marconi & Hunt 2003; McConnell & Ma 2013), and the central velocity dispersion σ (e.g. Ferrarese & Merritt 2000; Tremaine et al. 2002; Gültekin et al. 2009; McConnell & Ma 2013). Figure 1.6 shows the latest determination of the $M_\bullet - M_{\rm b}$ and $M_\bullet - \sigma$ relations from McConnell & Ma (2013).

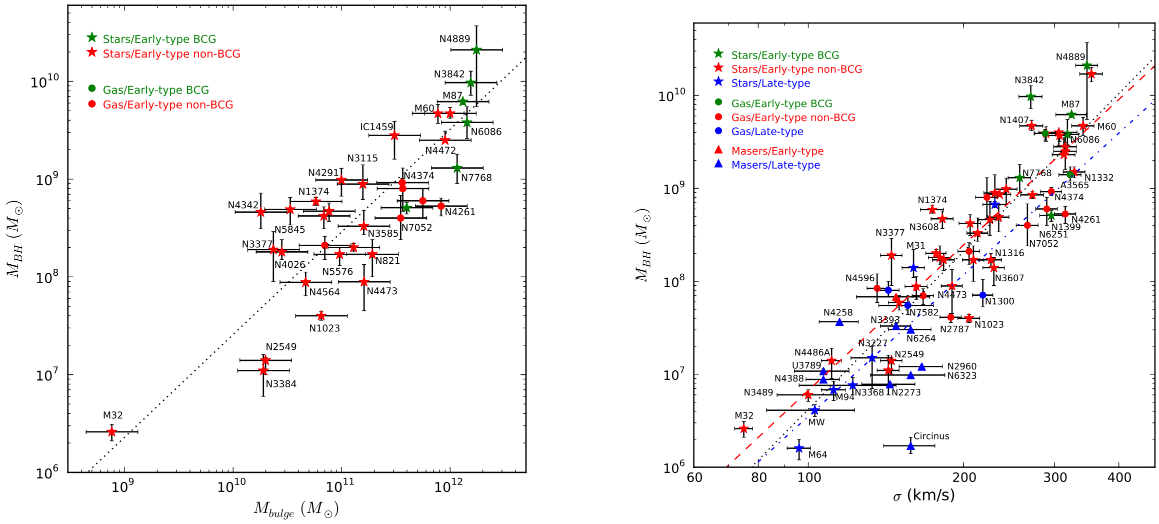

Figure 1.6 Left and right panels show the $M_\bullet - M_{\rm b}$ and the $M_\bullet - \sigma$ relations from McConnell & Ma (2013), respectively. Green, red and blue stars (circles) refer to star-based (gas-based) mass determinations in early-type galaxies at the centre of clusters, early-type galaxies not at the centre of clusters, and late-type galaxies, respectively. Blue and red triangles in the right panel refer to maser-based mass determinations in early- and late-type galaxies, respectively. The black dotted lines show the best fit over the whole samples, while the red dashed and the blue dot-dashed lines in the right panel refer to the best fit over the early- and late-type galaxy subsamples, respectively.

The indirect methods to measure supermassive black hole masses are based on the existence of the scaling relations with galactic properties, but more importantly the latter reveal a possible intimate connection between the evolution of supermassive black holes and their host galaxies. Although still debated (see e.g. Jahnke & Macciò 2011), it has been argued that these scaling relations might appear because of the feedback from the accretion luminosity on the surrounding galaxy. Specifically, the slope of the $M_\bullet - \sigma$ relation, $M_\bullet \propto \sigma^\alpha$, which observationally oscillates between 4 and 5, could potentially discriminate between scenarios where feedback operates in a momentum-conserving fashion ($\alpha = 4$; see e.g. King 2003) or energy-conserving fashion ($\alpha = 5$; see e.g. Silk & Rees 1998).

In the first case, the gas is supposed to be able to efficiently cool via Compton cooling; therefore, the main coupling between the gas and the central energy source is through momentum injection $\dot{p} = \dot{M}_{\rm out} v \sim L_{\rm Edd}/c$, where $\dot{M}_{\rm out}$ is the mass outflow rate swept up by the pushing radiation $\sim L_{\rm Edd}$ and v is the outflow velocity. In the second case, the cooling time is long enough that the luminosity from the central source heats up the gas, which in turn expands and pushes away the inner shells, roughly conserving their internal energy U, i.e. $0 \sim \dot{U} \simeq L_{\rm Edd} - 4\pi R^2 v_{\rm sh} \rho v_{\rm sh}^2$, where $R_{\rm sh}$ and $v_{\rm sh}$ are the radius and the velocity of the expanding gas, respectively. In both cases, assuming that the background galaxy is an isothermal sphere[11] characterised by the velocity dispersion σ and requiring that the outflow velocity is equal to the escape velocity, one can derive a relation of the form $M_\bullet \propto \sigma^{4-5}$ (with slightly better agreement for the momentum-conserving case). This means that the $M_\bullet - \sigma$ relation might represent the balance achieved when the feedback from the central supermassive black hole is just about to halt accretion by unbinding the galactic gas, which in turn represents the mass reservoir from which the central black hole is fed.

Despite the potential information hidden behind the $M_\bullet - \sigma$ relation about the coupling between the accretion luminosity and the host galaxy, the picture may be complicated by additional ways the black hole can interact with the surroundings. In fact, regardless whether the coupling is through momentum or energy, the description above assumes that accretion can produce a large radiative luminosity $L = \eta \dot{M} c^2 \sim L_{\rm Edd}$, i.e. that the AGN is highly accreting in a radiative-efficient mode with $\eta \sim 0.1$. In such conditions, the mechanisms previously described require that a fraction $\sim 0.05L$ couples to the gas to induce effective feedback, also dubbed as “quasar” mode. However, the AGN may exert feedback even when the radiative luminosity $L \ll L_{\rm Edd}$, i.e. when the black hole accretes in a radiative-inefficient mode with $\eta \ll 0.1$. Specifically, the feedback may be mechanical and come from radio jets (see Section 1.2.2), i.e. collimated outflows of relativistic particles accelerated close to black hole with a kinetic luminosity $L \sim 10^{44-45}$ erg s^{-1}. The jets inflate hot bubbles in the intergalactic gas on scales ~ 50 kpc, transferring momentum through repeated supersonic shocks as the gas is swept up, until the shocks become subsonic and the bubbles transfer energy to the gas through adiabatic expansion. Such mechanical feedback, also dubbed as “radio” mode, in combination with the quasar mode, might be responsible for the regulation of star formation in massive galaxies by heating the halo gas and preventing it to rapidly cool back onto the galaxy (Springel, Di Matteo & Hernquist, 2005; Croton et al., 2006; Fabian, 2012). Supermassive black holes are therefore a fundamental constituent of galaxies, because they do not just grow passively as they get fed by their hosts, but they might have at the same time an impact on the evolution of the galaxies hosting them, i.e. virtually on every present day massive galaxy.

---

[11] An isothermal sphere is a spherical model specified by the velocity dispersion σ and by the density $\rho = \sigma^2/(2\pi G r^2)$. The circular velocity is constant, $V_{\rm c}^2 = 2\sigma^2$.

1.2.4 Supermassive black hole formation

Despite our knowledge of AGNs, accretion physics, and cosmology, a fundamental question remains unanswered: how did supermassive black holes form? The answer to this question has to be searched in the realm of the early Universe physics (for a review, see Volonteri 2010). In fact, observations have unambiguously shown that supermassive black holes with masses $\gtrsim 10^9$ $\mathrm{M}_\odot$ power luminous AGNs at redshifts as high as $z \sim 6-7$ (Fan et al., 2006; Willott et al., 2010; Mortlock et al., 2011; Wu et al., 2015). Those objects are likely not the main population of supermassive black holes at high redshift because observational limitations prevents from detecting lower luminosity (and lower mass) AGNs. However, they pose a severe challenge for any theoretical model to explain their formation and growth. As an example, let us consider a supermassive black hole that can grow its mass at a constant fraction f_{Edd} of the Eddington rate $\dot{M}_{\mathrm{Edd}}$ inferred from the Eddington luminosity, equation (1.39), i.e. $\dot{M}_{\mathrm{Edd}} = L_{\mathrm{Edd}}/(\eta c^2)$. The evolution of the black hole mass $M_\bullet$ is then the solution of the differential equation $\dot{M}_\bullet = (1-\eta) f_{\mathrm{Edd}} \dot{M}_{\mathrm{Edd}}(M_\bullet)$, namely:

$$M_\bullet(t) = M_\bullet(t_0) e^{(t-t_0)/\tau}, \tag{1.40}$$

where $M_\bullet(t_0)$ is the initial mass at t_0, and:

$$\begin{aligned} \tau &\equiv f_{\mathrm{Edd}}^{-1}\, \tau_{\mathrm{Sal}} = f_{\mathrm{Edd}}^{-1} \frac{\eta}{1-\eta} \frac{c\kappa_{\mathrm{es}}}{4\pi G} \approx \\ &\approx 4.5 \times 10^7 \left(\frac{f_{\mathrm{Edd}}}{1}\right)^{-1} \left(\frac{\eta}{1-\eta}\right)_{\eta=0.1} \mathrm{yr}, \end{aligned} \tag{1.41}$$

where τ_{Sal}, the *Salpeter time*, is the timescale for accretion at the Eddington rate (Salpeter, 1964). Figure 1.7 shows $M_\bullet(t)$ at different f_{Edd}, with the requirement of matching the properties of J112001.48+064124.3, a massive AGN at $z \approx 7$ (Mortlock et al., 2011). This example clearly shows that it is very difficult to assemble the mass of those high-redshift AGNs in less than ~ 1 Gyr unless the initial black hole is either very massive at $z \gtrsim 20$ or it can accrete continuously at about the Eddington rate. However, in the latter case, feedback may also affect the growth process, with the tendency of limiting f_{Edd} to lower values.

Different models have been proposed to predict the initial population of black hole seeds. One possibility is to extend the concept of stellar mass black hole to the first generation of stars (PopIII). PopIII stars form from the collapse of metal-free gas clouds in haloes of about $\sim 10^6$ $\mathrm{M}_\odot$ at $z \sim 20-30$, where the only coolant is molecular hydrogen. They are expected to be more massive than present day stars; their mass spectrum can range from ~ 10 $\mathrm{M}_\odot$ to ~ 300 $\mathrm{M}_\odot$ (e.g. Bromm, Coppi & Larson 1999; Abel, Bryan & Norman 2002; Yoshida et al. 2006; but see also recent revisions of those masses towards lower values by e.g. Clark et al. 2011 and Greif et al. 2011). The most massive PopIII stars can then leave behind black holes with masses up to ~ 100 $\mathrm{M}_\odot$ at the end of their

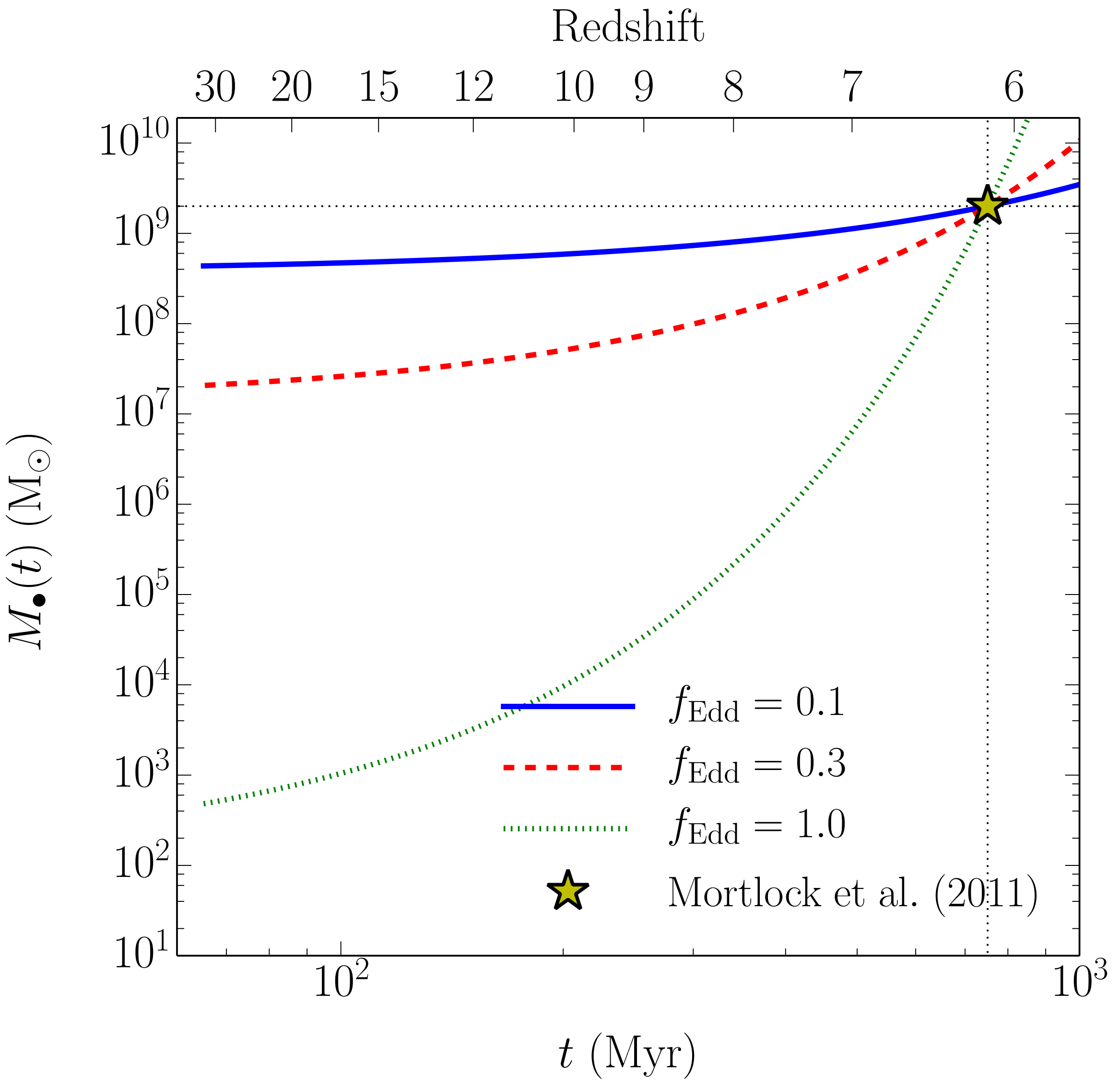

Figure 1.7 Evolution of supermassive black hole masses as a function of time at fixed fractions f_{Edd} of the Eddington rate. Solid blue, dashed red, and dotted green lines refer to $f_{\mathrm{Edd}} = 0.1$, 0.3, 1, respectively. All the lines reaches $M_\bullet = 2 \times 10^9$ at $z = 7.086$, i.e. the supermassive black hole in the quasar ULAS?J112001.48+064124.3 (Mortlock et al., 2011).

life, in analogy with massive stars in the local Universe. Those black holes might represent the initial supermassive black hole seeds (Madau & Rees, 2001; Tanaka & Haiman, 2009), though they would necessary grow at least at the Eddington rate in order to match the massive AGNs at $z \sim 6$ (see Figure 1.7).

A second possibility is that black hole seeds originate through dynamical processes within primordial nuclear star clusters, i.e. dense stellar clusters at the centre of protogalaxies (Quinlan & Shapiro, 1990; Devecchi & Volonteri, 2009; Devecchi et al., 2012). If massive stars can migrate to the centre of the cluster quickly enough, the density may become large enough that they undergo runaway mergers, i.e. a fast sequence of star-star mergers through which a very massive star ($\gtrsim$ 1000 $M_\odot$) might form. At the end its lifetime, the very massive star may core-collapse into a black hole up to $\sim$ 1000 $M_\odot$. Though less problematic than the PopIII case, such a seed would also require continuous accretion to grow.

A third possibility is the so called "direct collapse" scenario (e.g. Bromm & Loeb 2003; Begelman, Volonteri & Rees 2006; Lodato & Natarajan 2006; Dijkstra et al. 2008; Begelman & Shlosman 2009; Mayer et al. 2010; Latif et al. 2013; Mayer et al. 2015). Although there are different flavours for this approach, this scenario generally envisages a large mass of nearly pristine gas ($\sim 10^6 - 10^7$ $M_\odot$), promptly accumulated at the centre of a galaxy-size halo on (sub)parsec scales. However, this processes requires (i) to avoid fragmentation, and (ii) to effectively redistribute angular momentum to sustain the collapse on progressively smaller scales. Several possibilities have been discussed, such as the dissociation of H_2 molecules by Lyman-Werner ionising radiation coming from nearby, star-forming galaxies in order to avoid cooling and fragmentation (Ferrara & Loeb, 2013; Dijkstra, Ferrara & Mesinger, 2014). Another possibility is the onset of supersonic turbulence and the removal of angular momentum due to non-axisymmetric perturbations and gravitational torques during the collapse of the halo (Begelman & Shlosman, 2009; Choi, Shlosman & Begelman, 2013, 2015), or at the centre of major merger remnants between rare and massive galaxies at high redshift (Mayer et al., 2010, 2015). Then, a large fraction of the central cloud could rapidly ($< 10^6$ yr) collapse into a massive seed ($10^4 - 10^6$ $M_\odot$) directly at the centre of a galaxy at $z \sim 10 - 15$, either through general relativistic instabilities or via more exotic processes, such as the formation of a supermassive star (i.e. a radiation-dominated star with $\sim 10^5 - 10^6$ $M_\odot$), and then of quasi-star (i.e. a massive envelope sustained against its own self-gravity by the accretion luminosity of a central black hole; see Chapter 3 for additional details). Attractive as it is, this process is still far from being proved; nonetheless, it is the favoured model to explain the origin of the luminous high-redshift AGNs because values of $f_{\rm Edd} \sim 0.5$ would be enough to grow a massive seeds up to $\sim 10^9$ $M_\odot$ by $z \sim 6$.

1.2.5 Black hole binaries and gravitational waves

Supermassive black holes seem to be ubiquitous in massive galaxies (Ferrarese & Ford, 2005; Greene et al., 2010). This has relevant implications within the hierarchical paradigm of galaxy formation (see Section 1.1). In particular, the formation of supermassive black hole binaries looks a natural outcome when two galaxies merge (for a review, see Colpi & Dotti 2011).

This was first realised by Begelman, Blandford & Rees (1980), who described, in their seminal paper, the evolution of such a system through a sequence of three phases. Initially, (i) the pairing is dominated by dynamical friction, first of the galaxies against the dark matter background (at separations $\gtrsim$ 10 kpc), then of the galactic cores embedding the massive black holes against the evolving merger remnant (from separation $\lesssim 10$ kpc down to ~ 1 kpc). Dynamical friction is the response of a massive body M moving through a medium made of smaller bodies m with number density n ($\rho = mn$) at relative velocity V. During the motion, gravitational interactions between M and the background bodies tend to focus more m's behind M, forming a trailing wake that gravitationally pulls M against the direction of motion. It produces the acceleration (Chandrasekhar, 1943):

$$\mathbf{a}_{\rm DF} = -\frac{4\pi \ln\Lambda \, G^2 \rho M}{V^3} \, I(V) \, \mathbf{V}, \tag{1.42}$$

where $\ln\Lambda$ is called Coulomb logarithm and it is related to the typical minimum and maximum distances at which gravitational interactions between M and m's may occur, while $I(V) \sim 1$ is a function that depends on the mass and velocity distribution of the background matter. Then, (ii) the galactic cores merge and the orbital decay is driven by dynamical friction of the supermassive black holes themselves against the stellar and/or gaseous background, until they form a gravitationally bound binary. Finally, (iii) the binary shrinks via three-body scatterings off single stars or via gravitational torques from the surrounding gas, and eventually the two supermassive black holes merge together after emitting a burst of gravitational waves (Baker et al., 2006; Rezzolla, 2009). Although the details of this process in different conditions (i.e. when stars or gas dominate the mass of the surrounding environment) are still the subject of active research, assuming that this process would eventually take place is of momentous importance for the study of gravitational waves.

Gravitational waves are small perturbations $h_{\mu\nu}$ of the otherwise flat space-time metric $\eta_{\mu\nu}$ (for a more comprehensive discussion, see e.g. Misner, Thorne & Wheeler 1973). They satisfy the wave equation $\Box h_{\mu\nu} = 0$ in vacuum, where $\Box$ is the d'Alembert operator, and propagates at the speed of light c. They can be produced by moving systems that have a time-varying quadrupole moment[12]. In fact, black hole binaries (both stellar and supermassive black holes) are among the most luminous sources of gravitational waves, whose luminosity scales as:

$$L_{\rm GW} \sim 6.4 \times 10^{51} \left(\frac{\mu}{0.5 \ {\rm M}_\odot}\right)^2 \left(\frac{M}{2 \ {\rm M}_\odot}\right)^{4/3} \left(\frac{f}{100 \ {\rm Hz}}\right)^{10/3} \ {\rm erg \ s^{-1}}, \tag{1.43}$$

where μ and M are the reduced and total mass of the binary, respectively, and f is the orbital frequency. Despite the enormous amount of energy liberated, the detection of gravitational waves is very hard in practice. This is because the strain h, i.e. the amplitude of the wave, is very tiny:

$$h \sim 10^{-22} \left(\frac{\mu}{0.5 \ {\rm M}_\odot}\right) \left(\frac{M}{2 \ {\rm M}_\odot}\right)^{2/3} \left(\frac{f}{100 \ {\rm Hz}}\right)^{2/3} \left(\frac{D}{15 \ {\rm Mpc}}\right)^{-1}, \tag{1.44}$$

---

[12] The reason why a time-varying quadrupole moment is required is that, for a gravitational system on its centre of mass, the monopole does not vary because of conservation of mass, while the dipole is identically 0 because of the nonexistence of negative masses, and therefore the quadrupole is the first term that contributes to gravitational radiation.

where D is the distance of the binary. When a gravitational wave passes through some length L, the strain corresponds to the relative change of this length, i.e. $h \simeq \Delta L/L$. Although very small, this variation can be ultimately detected as a phase shift between two light rays in an interferometer. In fact, some Earth-based experiments are currently on-going to detect gravitational waves at high frequency (10^2-10^3 Hz) coming from stellar mass black hole binaries (e.g. LIGO Scientific Collaboration[13]; VIRGO Collaboration[14]). Recently, the LIGO Scientific Collaboration has announced the first detection of gravitational waves emitted by two black holes ~ 30 M$_\odot$ each at a distance ~ 400 Mpc (Abbott et al., 2016). On the other hand, the emission of gravitational waves from supermassive black hole binaries occurs at sub-Hz frequencies that require very long interferometric baselines to be detected; therefore, they will be the target of future space-based gravitational wave observatories (eLISA[15] will target supermassive black holes $M_\bullet \sim 10^{5-6}$ M$_\odot$ at $z \sim 2-5$; Amaro-Seoane et al. 2013), as well as ground-based detectors that exploit the signal from pulsars[16] to build an extremely long baseline (IPTA[17] will target supermassive black holes $M_\bullet \sim 10^{8-10}$ M$_\odot$ at $z \lesssim 1$; Hobbs et al. 2010).

---

[13] `http://www.ligo.org`

[14] `http://public.virgo-gw.eu/language/en/`

[15] `https://www.elisascience.org`

[16] Pulsars are spinning neutron stars with a collimated radio emission that periodically points toward the Earth. Their period is very stable over time, which makes them very precise cosmic clocks.

[17] `http://www.ipta4gw.org`

1.3 This Thesis

The purpose of this Thesis is to advance our understanding on different topics related to the evolution of high redshift galaxies, the formation of supermassive black holes, and the evolution of black hole binaries. Specifically, we will address the following issues:

- **Chap. 2**: we study the evolution of $z \sim 3$ galaxies ($M_\star \sim 10^8 - 10^{11}$ M$_\odot$) during the assembly of a group in a cosmological simulation, focusing on the processes that determine their morphological appearance and the assembly of their central bulges.
- **Chap. 3**: we explore the impact of outflows, driven by super-Eddington luminosities, as well as the role of rotation, which determines the angular momentum boundary conditions of the inner accretion region, on the evolution of quasi-stars – massive envelopes sustained by the accretion luminosity of a central black hole – postulated to be the cradle of massive black hole seeds.
- **Chap. 4**: we consider the evolution of massive black hole pairs in gaseous environments. Specifically, we focus in exploring the role of inhomogeneities of the surrounding gaseous background on the black hole dynamics.

2

The morphological evolution of $z \approx 3$ galaxies[1]

Abstract In this Chapter I will discuss the evolution of $z \sim 3$ galaxies with a broad range of stellar masses ($\log M_{\star}/\mathrm{M}_{\odot} \sim 8-11$) during the assembly of a galactic group in the Argo cosmological simulation. The galaxies show a diversity of realistic morphologies, including late-type/irregular disc galaxies with flat rotation curves, spheroid dominated early-type discs, and a massive elliptical galaxy, already established at $z \sim 3$. I will show that: (i) major mergers are the main trigger for the formation of bulges and the steepening of the circular velocity curves, and (ii) minor mergers and non-axisymmetric perturbations (stellar bars) drive the bulge growth in some cases. The results suggest that morphological transformations of high redshift galaxies of intermediate mass are likely triggered by processes similar to those at low redshift and result in an early build-up of the Hubble sequence.

---

[1] The material of this chapter is contained in and adapted from the following publication: **Fiacconi**, Feldman, & Mayer 2015, MNRAS, 446, 1957.

2.1 Introduction

The current morphology of moderately massive galaxies likely traces back to $z \sim 1-3$ (Conselice, Blackburne & Papovich, 2005; Buitrago et al., 2013) when the star formation rates of their progenitor galaxies were at their maximum (e.g., Leitner & Kravtsov 2011; Yang et al. 2013; Behroozi, Wechsler & Conroy 2013). By $z \sim 1$ the Hubble sequence was firmly in place (e.g., Brinchmann et al. 1998; Conselice, Blackburne & Papovich 2005; Ilbert et al. 2006; Oesch et al. 2010) with "peculiar" galaxies, ubiquitous at higher redshift, being a sub-dominant galaxy population (e.g., Driver et al. 1995; Scarlata et al. 2007; Mortlock et al. 2013). The large diversity of galaxy morphologies encoded in the Hubble sequence, ranging from ellipticals to late-type spirals and irregulars, poses one of the key questions of galaxy formation and evolution: Which mechanisms are responsible for the observed variety of galactic structural parameters, such as bulge-to-disc ratios, gas fractions, kinematics, and star formation rates (e.g. Hubble 1926; Sandage 1961; Dressler 1980; de Vaucouleurs et al. 1991; Roberts & Haynes 1994; Steinmetz & Navarro 2002; Gutiérrez et al. 2004; Driver et al. 2006; van der Wel 2008; Gavazzi et al. 2010, 2013)?

Galaxy interactions, such as minor and major galaxy mergers (Toomre & Toomre, 1972; Barnes, 1988; Naab & Burkert, 2003; Boylan-Kolchin, Ma & Quataert, 2005; Naab & Trujillo, 2006; Cox et al., 2006; Fiacconi et al., 2012; Hilz, Naab & Ostriker, 2013), galaxy harassment (Aguilar & White, 1985; Moore et al., 1996; Moore, Lake & Katz, 1998; Mastropietro et al., 2005), and tidal stirring of small galaxies by more massive ones (Mayer et al., 2001; Łokas et al., 2010; Kazantzidis, Łokas & Mayer, 2013), are often suggested as important transformation channels. In addition, secular processes induced by instabilities of the stellar component, such as bars, spiral arms or buckling/bending modes (Courteau, de Jong & Broeils, 1996; Kormendy & Kennicutt, 2004; Debattista et al., 2006), as well as gravitational instabilities of the gaseous disc (Bournaud et al., 2008; Dekel, Sari & Ceverino, 2009; Ceverino, Dekel & Bournaud, 2010; Bournaud et al., 2011, 2014; Moody et al., 2014) may result in a transformation from late- to early-type morphology.

Moreover, as the stellar component is largely built in-situ (e.g., Moster, Naab & White 2013; Behroozi, Wechsler & Conroy 2013), galaxy morphology is strongly linked to the baryonic cycle of gas accretion, consumption, and ejection. Mechanisms of the latter category may regulate or even completely remove the gas content of a given galaxy, thus transforming the object from a blue, star-forming system to a red, passively-evolving one by quenching star formation. Ram pressure stripping and strangulation (Gunn & Gott, 1972; Abadi, Moore & Bower, 1999), the transition from a cold to a hot mode accretion regime and gravitational quenching (Birnboim & Dekel, 2003; Dekel & Birnboim, 2006, 2008), outflows driven by stellar or AGN feedback (Scannapieco, Silk & Bouwens, 2005; Bundy et al., 2008; Davé, 2009; Oppenheimer et al., 2010; Teyssier et al., 2011; Dubois et al., 2013), and reduced gas accretion rates caused by cosmological starvation (Feldmann & Mayer, 2014) likely all contribute given the right circumstances.

There are hints that significant morphological evolution, including bulge growth via bar instabilities triggered by minor mergers and tidal interactions, may start quite early, at $z > 3$ (Guedes et al., 2013). However, works studying galaxy evolution at high-z mostly concentrate on galaxies as massive as the Milky Way today at $z \sim 2 - 3$, which may undergo significant evolution via violent disc instabilities triggered by high gas accretion rates (Agertz, Teyssier & Moore, 2009; Ceverino et al., 2012; Moody et al., 2014). These galaxies are the likely progenitors of some of the most massive present-day galaxies, such as those sitting at the centres of galaxy clusters and rich groups.

We extend the focus towards more typical galaxies with stellar masses between 10^8 and 10^{10} $\mathrm{M}_\odot$ at $z \sim 3 - 4$, which are likely progenitors of spiral galaxies and low- to intermediate-mass early-type galaxies today (Marchesini et al., 2009; Ilbert et al., 2013; van Dokkum et al., 2013). We primarily focus on the timescale and the mechanisms involved in generating their high redshift morphologies. In particular, we ask when the Hubble sequence of such typical galaxies was first established and whether galaxy evolution at high z is really different from low-z as models of violent disc instabilities would suggest.

Content of the Chapter In the following, we study the morphological evolution of $z \approx 3$ galaxies within the Argo cosmological zoom-in simulation, whose properties are discussed in Section 2.2. In Section 2.3 we present the results, based on synthetic RGB images, circular velocity curves, surface density profiles and specific angular momentum analysis. We find that a morphological sequence very similar to the Hubble Sequence, mostly among late-type galaxies, is already in place at $z \approx 3$. The processes responsible for morphological changes are similar to those operating at lower z, namely major mergers and secular evolution through the formation of bars, while violent disc instability does not seem to play a major role at these mass scales. We highlight potential issues that may affect the conclusions in Section 2.4, and we finally we discuss our main findings and conclude in Section 2.5.

2.2 Numerical simulations

2.2.1 Simulation code and initial conditions

The Argo simulation is a suite of zoom-in cosmological simulations first presented by Feldmann & Mayer (2014). The simulation is a follow-up of the $G2$ simulation described by Feldmann et al. (2010) and Feldmann, Carollo & Mayer (2011). Argo has a high-resolution region of ~ 3 comoving Mpc per edge centred on a dark matter halo of mass

$\sim 2 \times 10^{13}$ $\mathrm{M}_\odot$ at $z = 0$ inside a box of 123 comoving Mpc. This dark matter halo hosts at $z = 0$ a group dominated by a central, early-type, massive galaxy. It is worth to point out that this specific group resides and evolves in a slightly over-dense but still rather usual environment since the matter over-density $\delta \equiv \rho/\langle\rho\rangle - 1$ (where $\langle\rho\rangle$ is the average matter density) at $z = 0$ measured in $5\,h^{-1}$ Mpc around the group is $\delta = 1.4$ and the final halo mass is smaller than the exponential cutoff mass in the halo mass function at $z = 0$ and close to the halo mass ($\sim 10^{13}$ $\mathrm{M}_\odot$) corresponding to 1-σ fluctuations at $z = 0$ (e.g. Reed et al. 2003; Tinker et al. 2008; Watson et al. 2013).

The simulations were performed with the Tree/SPH N-body code GASOLINE (Wadsley, Stadel & Quinn, 2004) assuming a *Wilkinson Microwave Anisotropy Probe* 3-year cosmology (the same cosmology of the original $G2$ run) with $\Omega_{\mathrm{m},0} = 0.24$, $\Omega_{\Lambda,0} = 0.76$, $\Omega_{\mathrm{b},0} \simeq 0.04$, $h = 0.73$, $\sigma_8 = 0.76$ and $n = 0.96$ (Spergel et al., 2007). The force resolution is determined by the physical gravitational softenings $\epsilon_{\mathrm{DM}} = 250$ pc and $\epsilon_\star = 120$ pc for dark matter and baryonic particles, respectively. The softenings are kept fixed in physical coordinates after $z = 9$. The mass resolution of dark matter, gas and stellar particles is $m_{\mathrm{DM}} = 7.9 \times 10^5$ $\mathrm{M}_\odot$, $m_{\mathrm{g}} = 2.1 \times 10^4$ $\mathrm{M}_\odot$ and $m_\star = 7 \times 10^3$ $\mathrm{M}_\odot$, respectively. The initial conditions are made of 8,036,232 dark matter particles and 26,644,480 gas particles. The simulation reaches redshift $z \simeq 3.0$ with ~54,000,000 particles.

GASOLINE follows the dynamical evolution of baryonic and dark matter due to gravity and hydrodynamics, taking into account the underlying expansion of the Universe. It also implements sub-grid prescriptions for optically-thin radiative cooling, star formation, supernova feedback, mass loss from stellar winds and metal enrichment. Since Argo HR has the same mass and force resolution, we employ the same parameters of the Eris simulation (Guedes et al., 2011), a cosmological zoom-in simulation that successfully reproduced the properties of a Milky Way-like galaxy at $z = 0$. In particular, the gas is allowed to cool down to ~ 8000 K following a radiative cooling function for an optically-thin gas with primordial composition (Wadsley, Stadel & Quinn, 2004). Note that we do not employ metal-line cooling in this work in order to be consistent with the code setup adopted for the original Eris simulation (Guedes et al., 2011; Mayer, 2012). We included the effect of a uniform, redshift-dependent UV background (Haardt & Madau, 1996). The star formation is computed following Stinson et al. (2006): the local star formation rate is $\dot{\rho}_\star = \epsilon_{\mathrm{SF}}\rho_{\mathrm{g}}/t_{\mathrm{dyn}}$, where $t_{\mathrm{dyn}} = 1/\sqrt{G\rho_{\mathrm{g}}}$. The gas is eligible to form stars if (i) it is in an over-dense, converging flow, (ii) it is locally Jeans-unstable, (iii) the density $\rho_{\mathrm{g}} > \rho_{\mathrm{th}} = 5$ H cm^{-3} and (iv) the temperature $T_{\mathrm{g}} \leq 30000$ K. The star formation efficiency $\epsilon_{\mathrm{SF}} = 0.05$. Stellar particles are then stochastically spawned with initial mass $m_\star$ and represent a stellar population with a Kroupa, Tout & Gilmore (1993) initial mass function (also adopted by Kroupa 1998). Stars add energy, mass and metals back to the surrounding gas via supernova feedback and stellar winds. Both type Ia and type II supernovae inject 8×10^{50} erg each to the neighbouring gas particles. Type II supernovae originate from stars between 8 and 40 $M_\odot$ at the end of their life (determined from the parametrisation of Raiteri, Villata & Navarro 1996) and are modelled according to the analytical blast wave solution of McKee & Ostriker (1977). In particular, the thermal energy is distributed to gas particles inside the maximum radius that a blast wave can locally reach and the cooling of those particles is shut off for the time corresponding to the end of the snowplow phase of the blast wave. The cooling is not disabled for type Ia supernovae, the frequency of which is estimated from the binary fraction of Raiteri, Villata & Navarro (1996). They eject all the same amount of mass (1.4 $M_\odot$) and metals (0.63 $M_\odot$ of iron and 0.13 $M_\odot$ of oxygen; Stinson et al. 2006). Low-mass stars between 1 and 8 $M_\odot$ return part of their own mass and metals to the surrounding gas through winds according to Weidemann (1987). The typical fraction of mass lost by a particle because of stellar winds is ∼40 % (see Stinson et al. 2006 for additional details on the feedback model implementation).

2.2.2 Halo detection and selection

We use the AMIGA Halo Finder (AHF, Gill, Knebe & Gibson 2004; Knollmann & Knebe 2009) to identify dark matter halos and galaxies and to determine their properties, such as centres or virial radii. The virial radius is defined as the radius enclosing a mean matter density $\Delta(z)\,\rho_{\rm crit}(z)$, where $\rho_{\rm crit}(z)$ is the critical density to have a flat Universe, whereas $\Delta(z)$ is the z-dependent virial over-density defined by Bryan & Norman (1998). We selected dark matter halos at $z \simeq 3.4$ that (i) were more massive than 5×10^9 $M_\odot$ (corresponding to >6000 bound dark matter particles) and (ii) contained more than 10000 stellar particles. Our selection results in a sample of 22 halos with virial masses between 5.8×10^9 and 1.1×10^{12} $M_\odot$ hosting galaxies with stellar masses between $\sim 10^8$ $M_\odot$ and $\sim 10^{11}$ $M_\odot$. We checked that these galaxies have at least 100 star particles at $z \leq 10$ and typically >8000 star particles for $z \leq 4$. Then, we traced them backward and forward in time, matching the dark matter particles with the same ID inside each halo for all snapshots between $z = 10$ and $z = 3$. In particular, the main progenitor of an halo in the snapshot i is identified as the halo in the snapshot $i - 1$ that maximise the value $f_{\rm shared} = N_{\rm shared}/\sqrt{N_i N_{i-1}}$, where N_i and N_{i-1} are the number of dark matter particles of the two halos in snapshot i and $i-1$, respectively, and $N_{\rm shared}$ is the number of dark matter particles shared among the two halos. Since dark matter particles have the same mass, $f_{\rm shared}$ can be interpreted as the average mass contributed by a halo to another halo. Note also that $f_{\rm shared}$ and the procedure itself is independent on whether we proceed forward or backward in time with the matching.

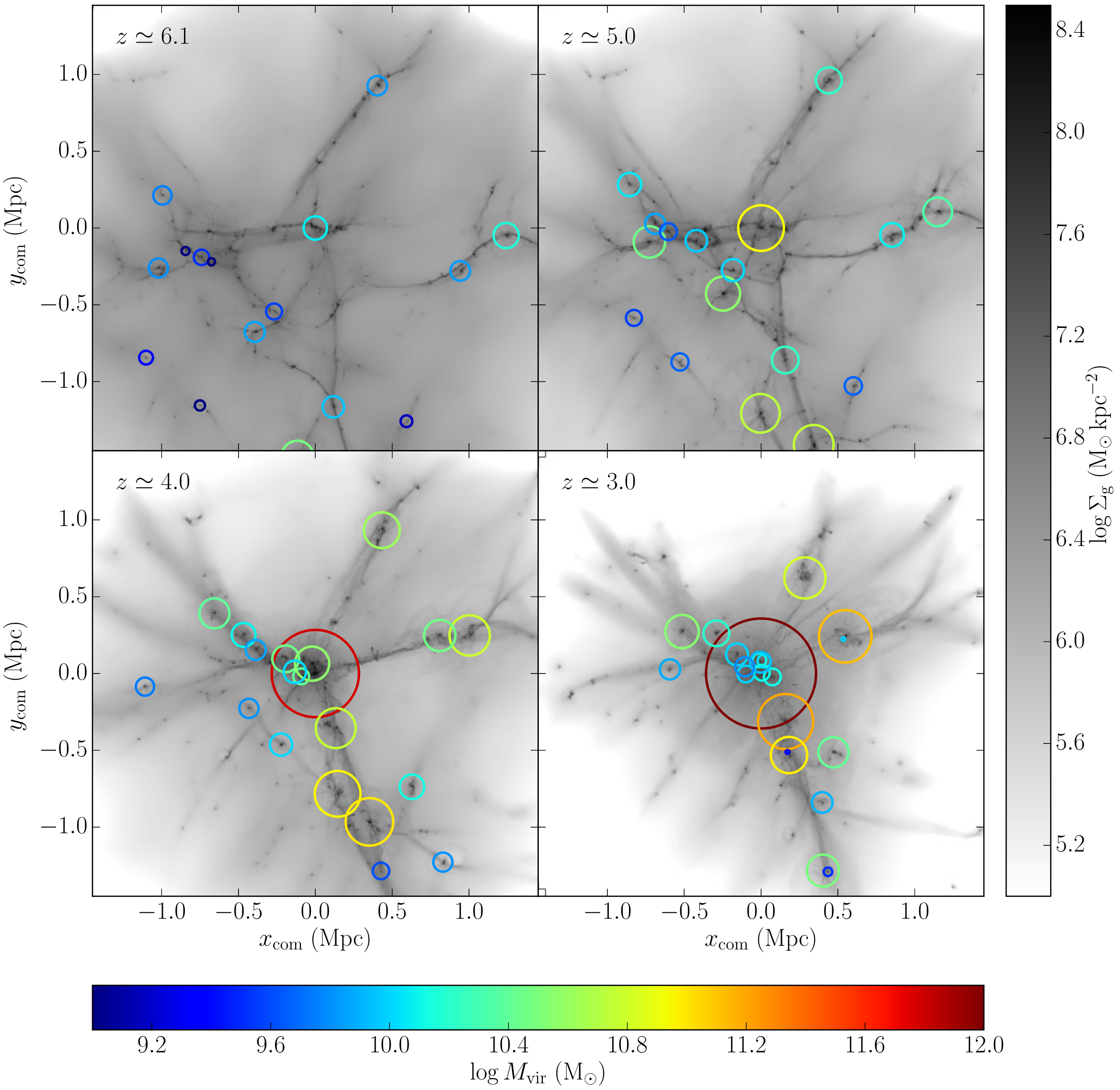

Figure 2.1 Gas surface-density maps for the high resolution region of the Argo simulation. The panels show snapshots at $z \simeq 6$, 5, 4, and 3 (from top left to bottom right, see legend). Circles shows the positions of the 22 galaxies in our sample. The circle sizes and colours indicate virial radii and virial masses of the parent dark matter halos of the selected galaxies. All panels are centred on the progenitor of the primary halo, i.e., on the progenitor of the halo harbouring the most massive galaxy at $z = 3$.

2.3 Results

2.3.1 Group evolution

Figure 2.1 shows the evolution of the high resolution region of the Argo simulation.

The main progenitor of the group-sized halo (the primary halo) forms at $z \gtrsim 7.5$ with a virial mass $\lesssim 10^{10}$ M$_\odot$, comparable to the mass of a few other halos in the high-resolution region of the simulation at that redshift. Between $z \sim 6$ and $z \sim 3.5$, the primary halo and its central galaxy grow quickly, outpacing the growth of other massive halos in the high-resolution region. The primary halo lies at the crossing and merging point of several main dark-matter filaments that focus the dark matter and gas flows toward it, with a typical cold, star-forming gas ($T \leq 30000$ K) inflow rate of ~ 10 M$_\odot$ yr^{-1} through the virial radius at $z \sim 5.5$. The cold gas penetrates inside $\sim 10\%$ of the virial radius, roughly where the central galaxy resides, sustaining an average star-formation rate of ~ 20 M$_\odot$ yr^{-1} within it. By $z \sim 4$, the central halo is by far the biggest halo with a virial mass $\sim 4 \times 10^{11}$ M$_\odot$ and a physical virial radius ~ 60 kpc. The main halo keeps growing up to $z \sim 3.5$ reaching $M_{\rm vir} \gtrsim 10^{12}$ M$_\odot$ due to increasing inflow rate of cold gas, but also due to accretion of small satellite galaxies that mostly form along the filaments and flow into the group, as shown in Figure 2.1. After $z \sim 3.5$, the main halo stops growing and enters the phase of "cosmological starvation" described by Feldmann & Mayer (2014).

Some of the other massive galaxies ($M_{\rm vir} \sim 10^{11}$ M$_\odot$ at $z \simeq 3$) form even before the central one around $z \sim 9 - 10$, but further away along the main filaments. They move slowly along the filaments and smoothly accrete cold gas at a typical rate $\sim 5 - 10$ M$_\odot$ yr^{-1}, similarly to the main halo, but almost steadily up to $z \sim 3$. They typically experience up to 1-2 relevant merger episodes with mass ratios between ~1:1 and ~1:5 in the redshift range $3 \leq z \leq 10$. Some of these galaxies also accrete a few gas-rich satellites with low stellar masses.

2.3.2 Galaxy morphology

Figure 2.2 and 2.3 show face-on and edge-on mock images of our sample of galaxies at $z \simeq 3.4$, respectively. The face-on view is defined as the view along the line of sight determined by the specific angular momentum of stars inside $\sim 3\%$ of the virial radius $r_{\rm vir}$, excluding any contributions from satellites. We assign a luminosity to each star particle from mass-to-light ratio tables[2] based on the isochrones and synthetic stellar population of the Padova group (Marigo et al., 2008; Girardi et al., 2010; Bressan et al., 2012). These tables span the stellar age interval from 4×10^6 to 12.6×10^9 yr, the metallicity interval from 5×10^{-3} to 1.6 $Z_\odot$ and assume the initial mass function from Kroupa, Tout & Gilmore (1993) and Kroupa (1998). We built the images using the computed emission in the I, V and B Bessell (1990) filters as the R, G and B channels, respectively, and we stretched the outcome with a power-law transformation (exponent $\gamma = 0.3$) for better visualisation.

The central galaxy has an early-type, elliptical morphology, with thin stellar shells

---

[2] Tables available at `http://stev.oapd.inaf.it/cgi-bin/cmd`.

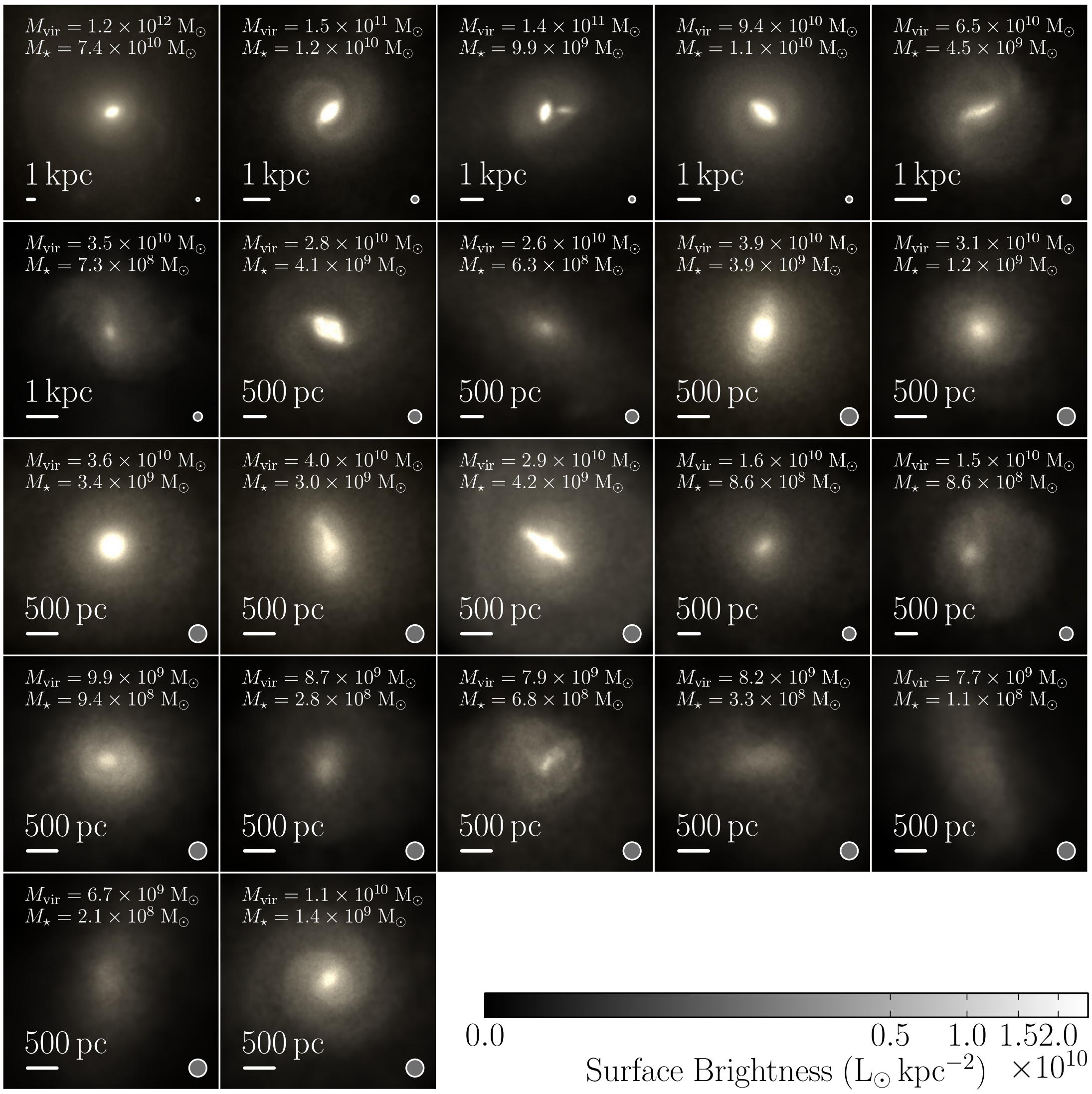

Figure 2.2 Face-on images of the 22 galaxies in our sample at $z \simeq 3.4$ in the B, V and I filter bands. Each panel encompasses roughly the inner 10% of the virial radius of each halo. The bottom left corner shows the scale of each image in proper pc. The stellar mass at $z \simeq 3$ and the virial mass at $z \simeq 3$ (or at the infall redshift for satellite galaxies at $z \simeq 3$) are provided at the top of each panel. The physical softening lengths of gas and star particles, $\epsilon_\star = 120$ pc, are indicated by the radius of the grey circle in the bottom right corner of each panel. Many galaxies show spiral arms or bars within extended stellar discs. Low mass galaxies often lack the concentrated central light concentration observed in more massive galaxies.

at a few physical kpc from the centre of the halo likely due to the tidal disruption of satellite galaxies on non-radial orbits (Hernquist & Quinn, 1988, 1989; Feldmann, Mayer & Carollo, 2008). The stellar mass at $z \sim 3$ within ~ 10 physical kpc is $\sim 7 \times 10^{10}$ M$_\odot$, with a total gas fraction (defined as the ratio between the gas mass and the sum of the stellar and gas mass) of $\sim 11\%$. The fraction of cold (≤ 30000 K), star-forming gas is only

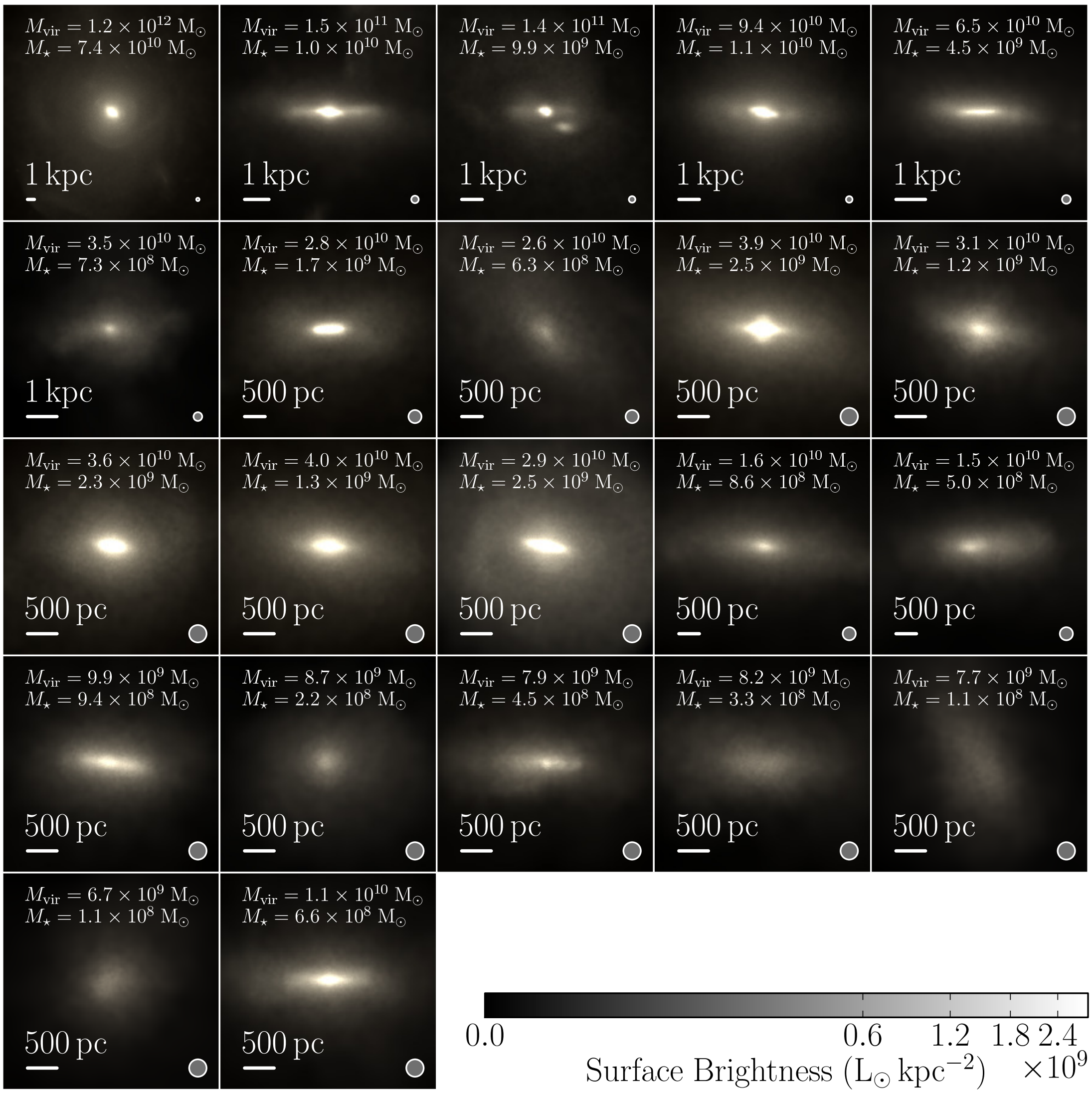

Figure 2.3 Same as Figure 2.2, but galaxies are shown edge-on. Most galaxies (but not the central galaxy of the primary halo) have a pronounced disc component. Lower mass galaxies have stellar discs that are thicker relative to the size of the galaxy.

$\sim 5\%$ and it is mainly distributed in a central disc with scale radius of ~ 450 physical pc. See (Feldmann & Mayer, 2014) for additional details on the central galaxy.

All the other galaxies exhibit a later-type morphology, ranging from extended, grand-design early-type spirals (likely Sa-Sb, from a visual inspection) to irregular, gas-dominated systems (Sm-Im, from a visual inspection). All these galaxies have specific star-formation rates ~ 2 Gyr^{-1} between $z \simeq 4$ and 3, consistent with the observed star formation sequence at $z \simeq 3.7$ by Lee et al. (2011). The most massive galaxies (virial masses $M_{\rm vir} \gtrsim$ a few 10^{10} M$_\odot$) typically have well-developed, grand-design spiral arms already at $z \gtrsim 3$. Those spiral arms are likely triggered by nearby satellites. However, numerical effects such as swing amplification induced by dark matter particles moving within the disc might also drive spiral structures. D'Onghia, Vogelsberger & Hernquist (2013) have shown that swing-amplified, multi-armed spirals can survive long after that the perturbation has faded away. This effect might be present in low mass systems that show flocculent spiral arms. The most massive disc galaxies often show a prominent bulge or a bar-like central distribution of stars, while lower mass galaxies are often pure discs. The discs have typical scale heights of $h \sim 400$ pc (a few softening lengths, estimated as the root mean square of the vertical displacement from the disc plane). The discs are thin (aspect ratios $\ll 1$) for the most massive and extended discs, while galaxies with $M_{\rm vir} \lesssim 10^{10}$ M$_\odot$ tend to develop thicker discs. This is likely a consequence of the stronger impact of stellar feedback in shallower potential wells. The stellar component of the lowest mass galaxies in our sample ($M_{\rm vir} < 10^{10}$ M$_\odot$, $M_\star \sim 10^8 - 10^9$ M$_\odot$) has a more irregular, although still flattened, morphology.

2.3.3 Circular velocity curves

Figure 2.4 shows the redshift evolution of the circular velocity profiles $V_{\rm c}(r) \equiv \sqrt{GM(r)/r}$, where $M(r)$ is the total mass inside the sphere of physical radius r, for the 22 galaxies in our sample. The circular velocity profiles evolve quickly from $z \lesssim 10$ to $z \simeq 3$ via the rapid accretion of dark matter. For instance, the normalisation of the asymptotic velocity $V_{\rm c}(r_{\rm vir})$ increases from ~ 20 km s^{-1} at $z = 10$ to $\gtrsim 150$ km s^{-1} at $z = 3$ for galaxies with $z = 3$ masses of $M_\star \sim 10^{10}$ M$_\odot$. At the same time, the shape of the circular velocity curves in the inner region is substantially affected by the distribution of baryons. In particular, we can clearly distinguish two kinds of circular velocity profiles up to $z \simeq 3$ in Figure 2.4. Some galaxies have flat or almost flat circular velocity curves (those galaxies will be dubbed as "flat" in the following), whereas other galaxies have a clear peak in the inner region (those galaxies will be dubbed as "peaked' in the following). In order to discriminate quantitatively between these two classes, we measured the ratio between the maximum circular velocity $V_{\rm c}(r_{\rm max})$ and the value $V_{\rm c}(r_{\rm h})$ at the radius $r_{\rm h}$ that encloses half of the *total* mass of the halo. We then classify a galaxy as "flat" or "peaked" at a given redshift if $V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h})$ is below or above the assumed threshold of 1.25. While the value 1.25 is to some degree arbitrary, our results are not strongly affected by the specific value that we assume. Moreover, we decided to use $V_{\rm c}(r_{\rm h})$ instead of $V_{\rm c}(r_{\rm vir})$ because we found this definition more robust against the specific way the halo is identified and against the exact procedure to determine whether a particle is bound to a halo or not.

Figure 2.4 suggests that being a "flat" or "peaked" galaxy depends weakly on the virial mass of the host halo. This is presented explicitly in the left panel of Figure 2.5, which shows the ratio $V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h})$ as a function of $M_{\rm vir}$ at three different redshifts. We compute the Spearman's rank correlation coefficient to assess more quantitatively the

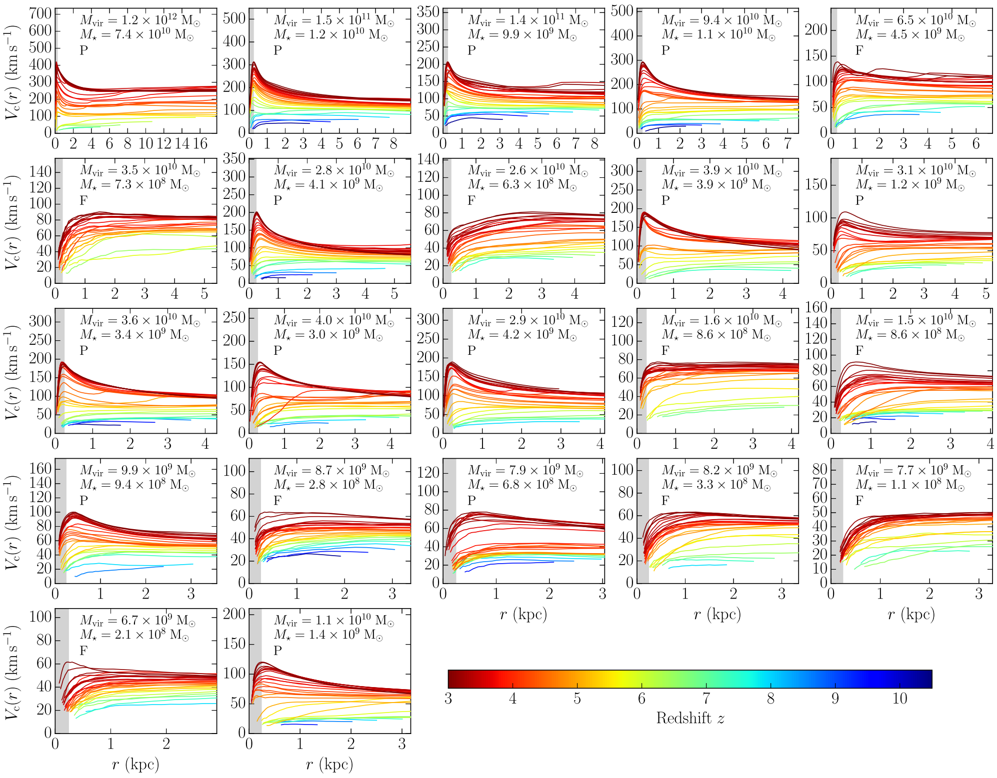

Figure 2.4 Circular velocity profiles for the 22 galaxies in our sample. Curves in each panel are color-coded according to redshift (spanning the range $3 \leq z < 11$). The profiles are shown out to 20% of maximally reached virial radius (radii are in physical units). The stellar mass at $z \simeq 3$ and the virial mass at $z \simeq 3$ (or at the infall redshift for satellite galaxies at $z \simeq 3$) are provided at the top of each panel. Each galaxy is flagged with "F" or "P" depending on whether it is classified as a "flat" or "peaked" galaxy at $z \sim 3$. The grey shaded areas correspond to 2 softening lengths, $r = 2\epsilon_\star$. Galaxies of lower (higher) stellar mass often show flat (peaked) circular velocity curves.

degree of correlation between $V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h})$ and $M_{\rm vir}$. We find values between ~ 0.3 and ~ 0.45 from $z \simeq 5$ to $z \simeq 3$, implying that the correlation is weak. This shows that having a "flat" or "peaked" circular velocity curve is neither strongly correlated with halo mass nor with the number of resolution elements (particles) per halo. On the other hand, the right panel of Figure 2.5 shows the same plot with $M_\star$ instead of $M_{\rm vir}$. In this case, the Spearman's coefficient ranges from ~ 0.4 to ~ 0.8 at the different redshifts, supporting the visual feeling that the ratio $V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h})$ correlates more strongly with $M_\star$ than with $M_{\rm vir}$.

We can also test whether the shape of the circular velocity profile depends on the age of the galaxy. In Figure 2.6 we plot the maximum over z of the ratio $V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h})$, the "peakedness" of the circular velocity profile, as a function of the formation redshift

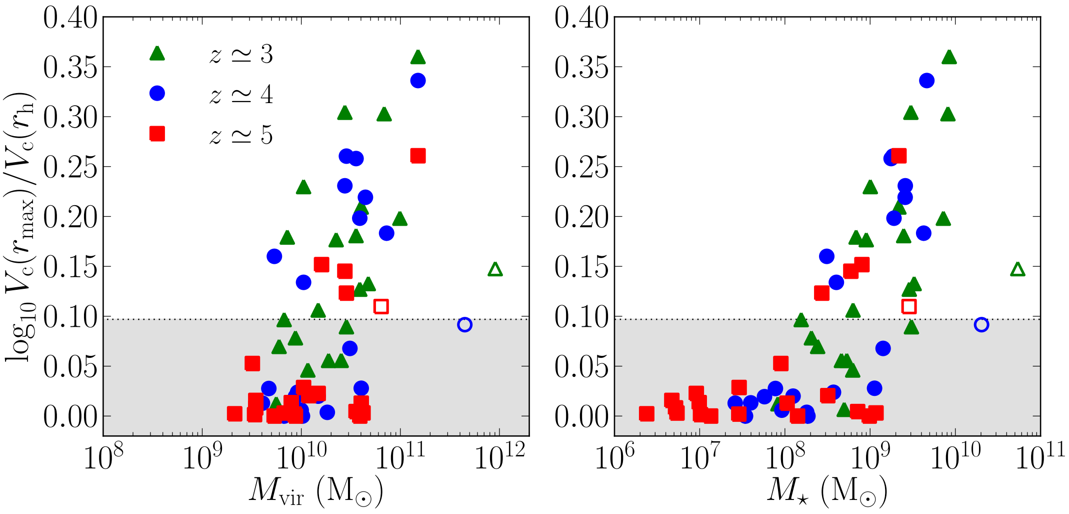

Figure 2.5 Peakedness of the circular velocity profile as function of virial mass M_{vir} (left panel) and stellar mass $M_\star$ (right panel) for the 22 galaxies in our sample. The peakedness is defined as $V_c(r_{max})/V_c(r_h)$. Here, r_{max} is the radius at which the circular velocity peaks and r_h is the radius containing half the virial mass. Red squares show the results for $z \simeq 5$, blue circles for $z \simeq 4$ and green triangles for $z \simeq 3$. The virial mass at infall time is used for satellite galaxies. Empty symbols denotes the central galaxy of the primary halo. The grey shaded area marks the region of "flat" systems, i.e., galaxies with $V_c(r_{max})/V_c(r_h) < 1.25$. The peakedness of the circular velocity profile is correlated with the stellar mass of intermediate-mass galaxies and only weakly with their halo masses.

z_{form}. The latter quantity is defined as the redshift at which the halo assembled at least $f_{form} = 5\%$ of its maximum mass at $z \geq 3$. The absence of a strong correlation between z_{form} and the peakedness of the circular velocity profile indicates that the shape of the mass distribution of our simulated galaxies is not merely related to the age of their parent halos. In other words, a larger value of z_{form} does not necessary imply that a galaxy is more "peaked". In particular, we find both "flat" and "peaked" galaxies with any $z_{form} > 6$. Our results do not change qualitatively if we define z_{form} based on a different value for f_{form}. Our finding suggests that the evolution of the circular velocity profile is not simply dictated by secular evolution.

2.3.4 The connection between V_c and morphology

The circular velocity profile $V_c(r)$ depends primarily on the mass distribution within r. The mass distribution within 10% of r_{vir} is dominated to a large extent by the baryonic (and in particular the stellar) mass. Therefore, a connection between the central distribution of stars and the shape of the circular velocity curve is expected. Figure 2.7 shows the stellar surface density profiles for the galaxies in our sample at $z \simeq 3$. The projection is face-on with respect to the axis given by the specific angular momentum of stars inside $0.03r_{vir}$. We perform a "partial" profile decomposition based on the surface

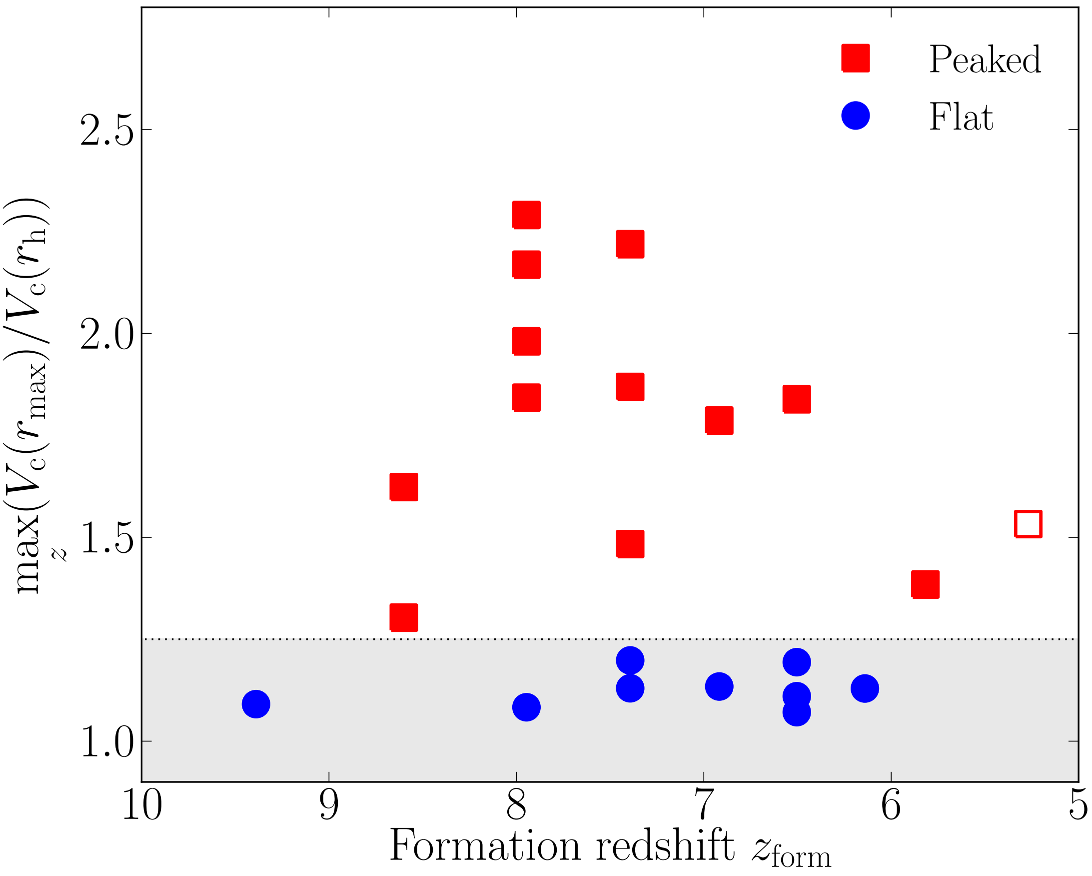

Figure 2.6 Maximum value of $V_c(r_{max})/V_c(r_h)$ over the redshift range $3 \leq z < 11$ as a function of the formation redshift z_{form}. Blue circles and red squares refer to "flat" and "peaked" galaxies, respectively. The empty symbol denotes the central galaxy. The grey shaded area marks the threshold $V_c(r_{max})/V_c(r_h) = 1.25$. The peakedness of the circular velocity profile is not strongly correlated with formation redshift.

density profiles. Instead of fitting the superposition of a bulge and a disc profile at the same time, we fit an exponential profile at large radii ($r \gtrsim 0.05 r_{vir}$) to model a stellar disc, assuming that the bulge contribution to the surface density at these radii is negligible. We then infer the bulge mass by subtracting the mass of the fitted exponential profile from the overall stellar mass within $0.1 r_{vir}$. We choose this simplified procedure because many galaxies show complex nuclear morphology (e.g. stellar bars) that make difficult a full profile decomposition and we only aim at a single-parameter morphological characterisation. Note that our determination of the bulge mass includes the mass of an eventual bar, as happens for many of the more massive galaxies of our sample (see Figure 2.2). Moreover, our method bases the estimates of both the bulge and disc masses on the surface density profiles. This might introduce some biases in the inferred B/T and B/D ratios with respect to what would be observationally determined via decomposition of surface

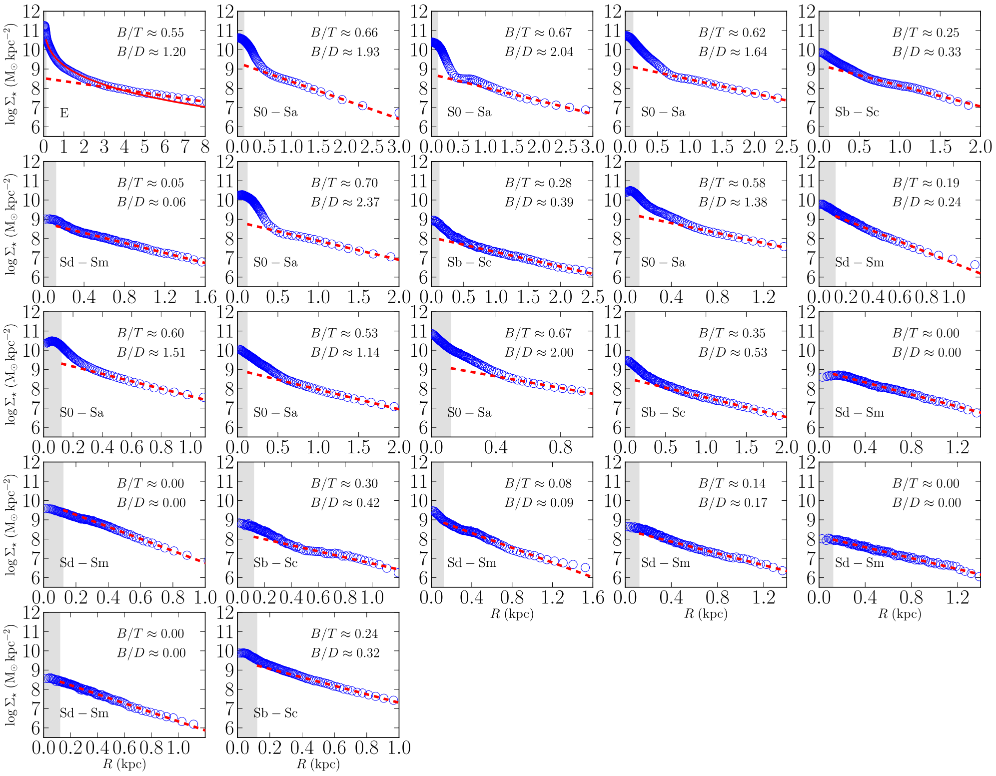

Figure 2.7 Stellar surface density profiles for the 22 galaxies in our sample at $z \simeq 3$. Blue circles show the data from the simulation, while red-dashed lines indicate the result of exponential fits to the stellar surface density at $r = 0.05 - 0.1r_{\rm vir}$. The red continuous line in the first panel shows the de Vaucouleurs fit for the central galaxy used in Section 2.3.5. We show the inferred bulge-to-total (B/T) and bulge-to-disc (B/D) ratios at the top right of each panel. We infer Hubble types based on the B/T ratio and the presence of a stellar disc as described in the text. The grey shaded areas correspond to a projected radius $\leq \epsilon_\star$. Galaxies with higher stellar mass show a larger central excess, indicative of a larger B/T ratio and an earlier Hubble type.

brightness profiles. In particular, we expect that our estimates are compatible with those obtained from near-infrared observations (e.g. in the H or K band), but they would likely overestimate the bulge component when compared with photometric measurements from bluer bands (e.g. B or V bands). Finally, we do not attempt a kinematic decomposition between the bulge and the disc, which is also known to produce systematically different estimates of B/D ratios (Scannapieco et al., 2010).

Based on the values of B/T (defined as the ratio of bulge mass to overall stellar mass), we assign to each galaxy a Hubble type: (i) E/S0-Sa for galaxies with $B/T \geq 0.4$, (ii) Sb-Sc for galaxies with $0.4 > B/T \geq 0.2$, and (iii) Sd-Sm for galaxies with $B/T < 0.2$. We distinguish between E and S0-Sa morphology based on the visual confirmation of an extended disc component. Moreover, some low mass galaxies have $B/T \approx 0$, even though

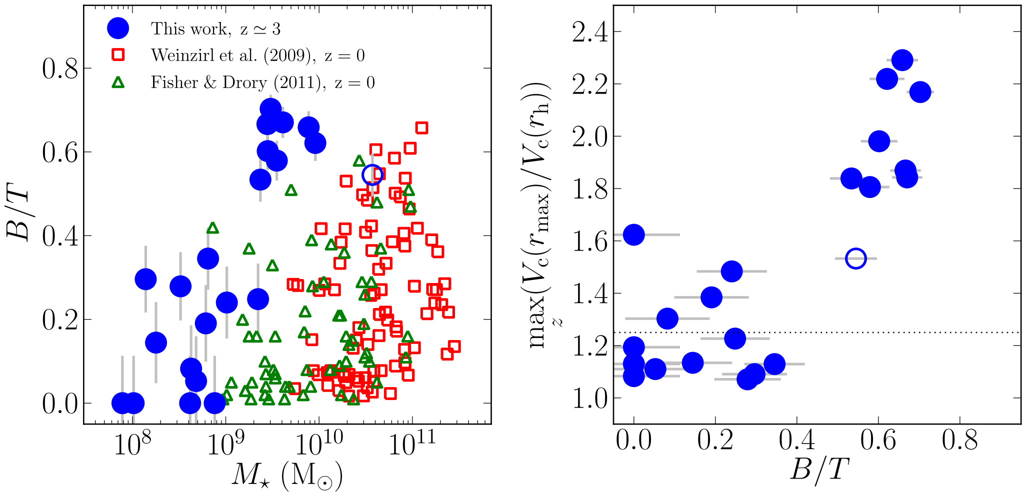

Figure 2.8 B/T ratio vs stellar mass at $z \simeq 3$ (left panel) and maximum peakedness of the circular velocity profile vs B/T at $z \simeq 3$ (right panel). Blue circles show our galaxy sample, red squares are observed B/T ratios by Weinzirl et al. (2009) and green triangles are from Fisher & Drory (2011). The horizontal dotted line in the right panel marks the threshold $V_c(r_{\max})/V_c(r_h) = 1.25$ that divides between "flat" and "peaked" galaxies. We derive errorbars assuming 10% uncertainty in the determination of both disc and total mass. The empty circle denotes the central galaxy in each panel. The peakedness of the circular velocity profile is a good indicator of the B/T ratio of galaxies. The morphology of the galaxies in our sample ranges from effectively bulge-less disc galaxies to spheroidal-dominated galaxies.

they might not appear as disc-dominated in Figure 2.2 and 2.3. Nevertheless, they have a flattened stellar component and a high gas fraction distributed on a rotationally supported disc that qualify them as late-type, disc-like systems. Therefore, although the face value $B/T \approx 0$ is a byproduct of our simplified procedure, this is still consistent with the typically observed $B/T \sim 0.1 - 0.2$ for Sd and later-type galaxies (e.g. Weinzirl et al. 2009) as long as we consider a reasonable uncertainty $\sigma \simeq 0.15$ on our B/T estimates (we obtain σ assuming 10% uncertainty on both the bulge and disc masses, see below). Although the exact B/T values for the subdivisions are somewhat arbitrary, they are chosen in accordance with the morphological classification at $z = 0$ (Scodeggio et al., 2002; Graham & Worley, 2008; Weinzirl et al., 2009).

Comparing Figure 2.4 and 2.7 suggests that the most massive and "peaked" galaxies have higher values of the B/T ratio. We show this result more explicitly in Figure 2.8, where we plot B/T as function of stellar mass and of the peakedness of the circular velocity. The errorbars on B/T are derived assuming an uncertainty of 10% for both the disc and total mass measurements. The trend between B/T and the stellar masses $M_\star$ is similar to the one found by Weinzirl et al. (2009), although their results are based on a sample of local spiral galaxies ($z = 0$) of various morphological type (from S0 to Sm) and with larger stellar masses $\gtrsim 10^{10}$ $M_\odot$ (determined from H band photometry). We note that the B/T values drawn from Weinzirl et al. (2009) include the mass fraction of the reported bar component. This is reasonable because we do not separate the non-disc component of our galaxies into a bar and a true bulge. We also compare our results with the data of Fisher & Drory (2011). Their sample at $z = 0$ partially overlaps in mass with ours and shows a comparable range of B/T.

As expected, the values of B/T measured from the simulation are highly correlated with the maximum of $V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h})$ over redshift. Hence, the evolution of the shape of the circular velocity curve is intimately connected with the morphological evolution, such that "peaked" galaxies are associated with more bulge-dominated galaxies. In other words, we can study the morphological evolution of galaxies by analysing the history of their circular velocity profiles. Note, however, that we do not distinguish between pure bulges and pseudo-bulges as we do not perform a complete profile decomposition and we do not identify bulges based on their Sérsic indexes (Kormendy & Kennicutt, 2004).

2.3.5 Morphology and the angular momentum content

Romanowsky & Fall (2012) and Fall & Romanowsky (2013) recently revised the early findings of Fall (1983) that the stellar specific angular momentum $j_\star$ of galaxies correlates with the stellar mass $M_\star$ and that the correlation itself varies with morphological type. We compare our simulation with these results in Figure 2.9. The upper panel of this figure reproduces Figure 2 of Fall & Romanowsky (2013), which shows the correlation between $j_\star$ and $M_\star$ for the disc component alone of (early- and late-type) spiral galaxies and for elliptical galaxies. They estimate the disc-only contribution to $j_\star$, $j_{\star,\rm d}$, as $j_{\star,\rm d} = 2V_{\rm rot,s}R_{\rm d}/\sin i$, where $R_{\rm d}$ is the scale radius of the exponential disc, $V_{\rm rot,s}$ is the asymptotic rotation velocity observed at large radii, and $\sin^{-1} i$ is a deprojection factor (Romanowsky & Fall, 2012). $j_{\star,\rm d}$ is plotted in Figure 2.9 against the disc mass $M_{\star,\rm d} = (1 - B/T)M_\star$. They treat elliptical galaxies as "pure bulges" and then estimate $j_\star$ as $j_{\star,\rm b} = Ck_nR_{\rm eff}V_{\rm rot,s}$, where C is an inclination correction factor (1.21 for lenticulars and 1.65 for ellipticals), k_n is a coefficient depending on the Sérsic index n, $R_{\rm eff}$ is the projected effective radius, and $V_{\rm rot,s}$ is the rotation curve evaluated at $\approx 2R_{\rm eff}$ (Romanowsky & Fall, 2012).

Mimicking the above procedure, we measure $j_{\star,\rm b}$ and $j_{\star,\rm d}$ for the central galaxy and the remaining galaxies of our sample, respectively. First, we derive the gas rotation curves in the plane of the gaseous disc for all our galaxies so that we do not need any deprojection factor. The gas rotation curve of the central galaxy is obtained in the plane of the central gaseous disc (see Section 2.3.2). In particular, we select a slice of ~ 500 pc centred in the plane of the disc (mimicking the action of a slit during the observation of an edge-on disc galaxy) and we measure the gas velocity as a function of radius projected along 20 random, polar line of sights (i.e. line of sights in the disc plane pointing toward the disc centre).

We finally average all the obtained rotation velocity curves[3]. We then estimate $j_{\star,\mathrm{d}}$ of all but the central galaxy as $j_{\star,\mathrm{d}} = 2R_\mathrm{d}V_\mathrm{s}$, where V_s is the asymptotic value of the gas rotation curve (typically similar, but slightly below, the asymptotic value of the circular velocity curve due to non-circular motions) and R_d is given by the exponential fits on the surface density profiles (Figure 2.7). Although we can decompose the surface density profile of the central galaxy in an exponential disc component and a remaining bulge component, a visual inspection clearly shows the lack of an extended disc. Therefore, we treat this galaxy as an elliptical, "pure bulge" galaxy in the procedure described above. We fit the surface density profile with a de Vaucouleurs profile (see Figure 2.7) to determine R_eff and we compute $j_{\star,\mathrm{b}} = 2.29V_\mathrm{s}R_\mathrm{eff}$, where $k_4 = 2.29$ and V_s is the rotation velocity at $\approx 2R_\mathrm{eff}$.

We also estimate the total specific angular momentum following Romanowsky & Fall (2012) as $j_\star = (1 - B/T)j_{\star,\mathrm{d}} + (B/T)j_{\star,\mathrm{b}}$, deriving the effective radii of bulges from the residual surface density profile after removing the exponential disc components. Then, we compare these estimates with direct measurements of the stellar angular momentum of the galaxies, i.e. $j_\star = \sum \mathbf{r} \times m_\star\mathbf{v}$, where $\mathbf{r}$ and $\mathbf{v}$ are computed in the reference frame of the centre of mass of the galaxy. We find values in fair agreement within a factor of ~ 2. We show our direct measurements in the lower panel of Figure 2.9, which reproduces Figure 3 of Fall & Romanowsky (2013). As expected, "flat" galaxies tend to have slightly higher $j_\star$ than "peaked" galaxies with comparable stellar mass. This trend is similar to the trend with morphology revealed by the observational data.

Both the upper and the lower panel of Figure 2.9 show that our data sit on a relation similar to the observed one, with a slope $\alpha \sim 0.6$ for the disc-only components (upper panel). However, the normalisation is different from that at $z = 0$. We argue that this discrepancy is due to the redshift evolution of this relation. This can be qualitatively understood, at least for the disc-only components, using analytical models for galaxy formation. In particular, following the simplest treatment of Mo, Mao & White (1998), we can model a galaxy as an exponential, non-self-gravitating disc embedded in an isothermal halo, from which we derive the expected specific angular momentum:

$$j_{\star,\mathrm{d}} = \frac{(2G)^{2/3}\lambda}{\Delta^{1/6}(z)H^{1/3}(z)} \left(\frac{f_j}{f_\star^{5/3}}\right) M_\star^{2/3}, \tag{2.1}$$

where λ is the spin parameter of the host halo, $\Delta(z)$ is the z-dependent virial overdensity, $H(z)$ is the Hubble parameter, f_j is the fraction of halo angular momentum retained by the disc, and $f_\star = M_\star/M_\mathrm{h}$ is the ratio between the disc and the halo mass. We can then estimate the redshift evolution between $z = 3$ and $z = 0$ of $j_{\star,\mathrm{d}}$ at fixed $M_\star$ as:

$$\left.\frac{j_{\star,\mathrm{d}}(z=0)}{j_{\star,\mathrm{d}}(z=3)}\right|_{M_\star} = \left(\frac{\Delta(z=0)}{\Delta(z=3)}\right)^{-1/6} \left(\frac{H(z=0)}{H(z=3)}\right)^{-1/3} \sim 2. \tag{2.2}$$

---

[3] Since we measure the rotation velocity as the velocity projected on the line of sight in the plane of the disc, we do not need any deprojection factor because in our case $i = 90°$ and $\sin^{-1} i = 1$.

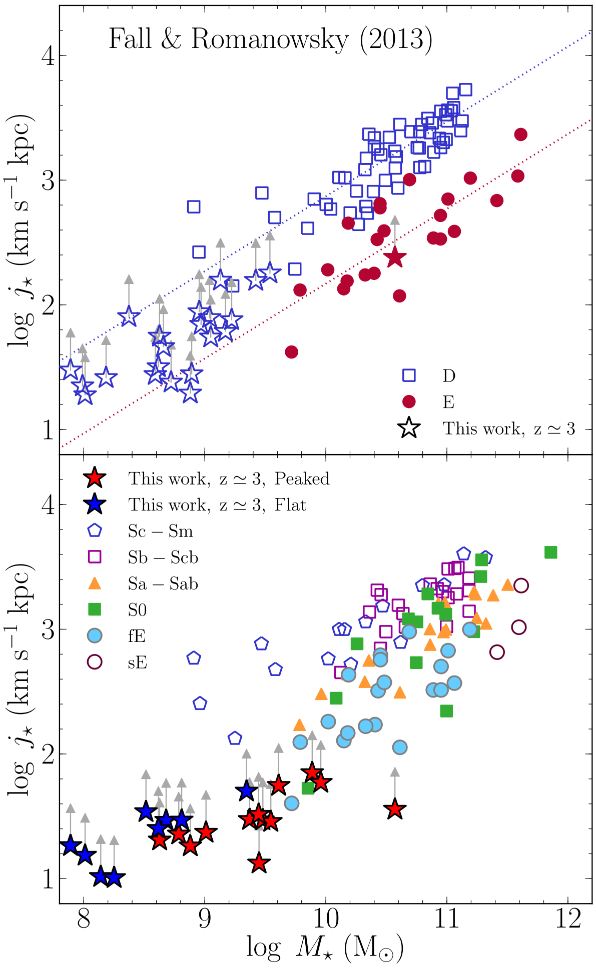

Figure 2.9 Stellar specific angular momentum $j_\star$ as a function of $M_\star$. Upper panel: blue, empty symbols denote disc-only components (i.e. $j_{\star,\rm d}$ vs. $M_{\star,\rm d}$, see the text for details), whereas red, filled symbols denotes elliptical galaxies made by bulge only (i.e. $j_{\star,\rm b}$ vs. $M_\star$). Empty squares and filled dots are observational data from Fall & Romanowsky (2013). Stars are the data from our galaxy sample at $z \simeq 3$. The upper and lower dotted lines represent the fits to the disc-only and bulge only relations, respectively (see Fall & Romanowsky 2013 for details). Lower panel: stars denote our direct measure of the total $j_\star$ as a function of total stellar mass $M_\star$ (see the text for details). Red stars mark “peaked” galaxies while blue stars mark “flat” galaxies. The other symbols are observational data from Fall & Romanowsky (2013) divided by morphological type according to the legend. In both panel, grey vertical arrows show an increment of $j_\star$ by a factor of 2 at constant $M_\star$, consistent with the evolution with redshift predicted by Equation (2.2).

This calculation does not account for the presence of a bulge or for additional processes that might change $j_\star$ (e.g. mergers, feedback, etc.). Nonetheless, it captures qualitatively the necessity of a redshift evolution for $M_\star - j_\star$ relation and is enough to explain and to cure the discrepancy between our $z \simeq 3$ data and the observations of nearby galaxies in the upper panel of Figure 2.9. The lower panel, instead, require a more detailed treatment of the build-up of galactic discs, although our simple correction is already enough to bring the simulated galaxies within the larger scatter expected at lower masses. Note that f_j

and $f_\star$ can also contribute to the redshift evolution of the relation in an $M_\star$-dependent fashion.

2.3.6 What triggers the difference between "peaked" and "flat" galaxies?

Our sample of 22 galaxies at $z \geq 3$ can be divided in two subclasses, "flat" and "peaked" galaxies, as discussed above. Out of 22 galaxies, 9 never reach the threshold $V_\mathrm{c}(r_\mathrm{max})/V_\mathrm{c}(r_\mathrm{h}) = 1.25$ and they are thus classified as "flat". The remaining 13 galaxies become "peaked" at redshift $z \geq 3$, most of them at $z > 4$.

All the galaxies that belong to the "flat" group, which represents $\sim 40\%$ of our galaxy sample, have $V_\mathrm{c}(r_\mathrm{max})/V_\mathrm{c}(r_\mathrm{h}) \simeq 1$ almost all the time. They are isolated galaxies that live in peripheral dark matter filaments. They do not approach the central galaxy for more than 2 virial radii of the primary halo up to $z \simeq 3$ and they do not experience any galaxy merger with mass ratios $\geq 1:5-1:6$. Their star formation rates are slowly rising for $z > 3$ and typically always below 1 $\mathrm{M}_\odot$ yr^{-1}, apart from short, burst-like periods when they reach up to ~ 2 $\mathrm{M}_\odot$ yr^{-1}.

On the other hand, the value of $V_\mathrm{c}(r_\mathrm{max})/V_\mathrm{c}(r_\mathrm{h})$ for 8 out of 13 "peaked" galaxies, which represent $\sim 60\%$ of the "peaked" subsample and $\sim 36\%$ of our entire sample of galaxies, suddenly changes from $V_\mathrm{c}(r_\mathrm{max})/V_\mathrm{c}(r_\mathrm{h}) \simeq 1$ to $V_\mathrm{c}(r_\mathrm{max})/V_\mathrm{c}(r_\mathrm{h}) > 1.25$. This change is clearly associated with a burst in star formation that can reach $\gtrsim 10$ $\mathrm{M}_\odot$ yr^{-1}. The starburst is triggered in each case by a major merger with a mass ratio $\geq 1:4$ for both M_vir and for $M_\star$, except for a couple of cases where the stellar mass ratio only is $< 1:4$.

We demonstrate this finding in Figure 2.10, where we follow the evolution of two example galaxies of our sample from before to after they merge with a lower mass companion galaxy. In both examples, the peak circular velocity increases dramatically during the galaxy merger and the associated starbursting phase. During that time the circular velocity profile transforms from flat to centrally peaked. The radius at which the circular velocity reaches its maximum moves inwards from $r_\mathrm{h} \sim r_\mathrm{vir}/2$ ($\sim 5-10$ kpc for the two example galaxies) to less than a kpc. These major mergers occur at different redshifts and in different environments. The two galaxies in the first example have a total mass ratio $q \simeq 1:2.9$ and merge around $z \simeq 6.7$ moving along a filament far from the primary halo. The second example involves two galaxies with a total mass ratio $q \simeq 1:3.6$. They merge around $z \sim 4.5$ in the vicinity (at $\sim 2r_\mathrm{vir}$) of the primary halo, i.e., in a mildly overdense environment.

The peaked circular velocity curves in the remaining five galaxies do not originate in major mergers. These five cases represent $\sim 38\%$ of the "peaked" subsample and $\sim 22\%$ of the total sample. The first of these anomalous galaxies is the central galaxy of

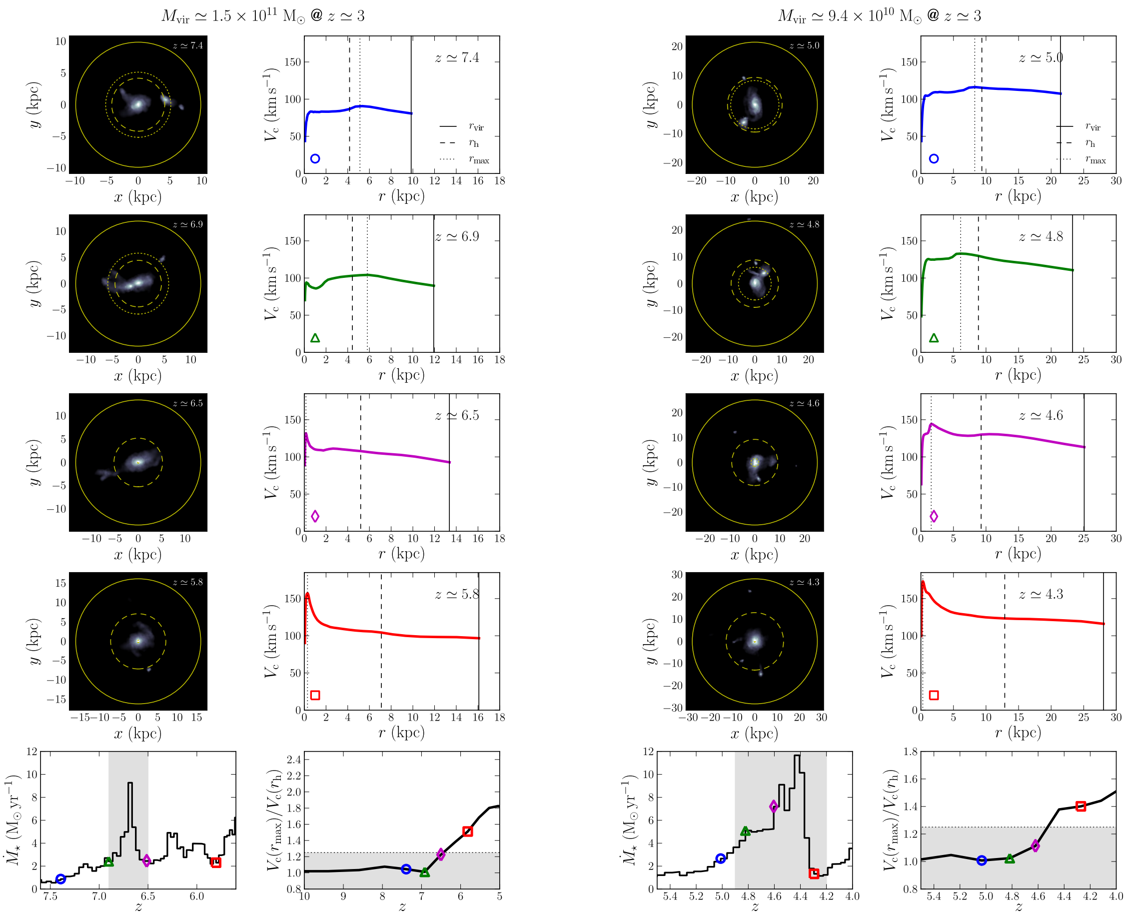

Figure 2.10 Evolution of the circular velocity curve for two example galaxies that obtain a peaked circular velocity profile via a major merger. The galaxy on the left (right) corresponds to the second (fourth) galaxy in the first row of Figure 2.4. For both example galaxies, we show the projected stellar density and the associated $V_{\rm c}(r)$ curve at four times during the galaxy interaction/merger (first four rows). Continuous, dashed and dotted lines mark the position of the virial radius, the position of the radius containing half of the total mass and the position of the radius corresponding to the peak of $V_{\rm c}$, respectively. The last row shows the evolution of the star formation rate within the central kpc (first panel) and the evolution of $V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h})$ (second panel). In the bottom row panels blue circles, green triangles, magenta diamonds and red squares mark the redshifts of the corresponding four snapshots. In both cases the peakedness of the circular velocity profile increases significantly over the course of the major merger.

the primary halo. This galaxy undergoes a large number of repeated mergers (but not major mergers) and galaxy interactions with orbiting satellite galaxies of lower mass. Its circular velocity profile starts to get strongly peaked only at relatively late times $z < 4$. In fact, at $z \simeq 4$ the circular velocity curve of this galaxy would be classified as "flat" ($V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h}) \sim 1.2$ at that time). The second of the anomalous galaxies experiences a late ($z \sim 3.6$) minor merger that is responsible for increasing $V_{\rm c}(r_{\rm max})/V_{\rm c}(r_{\rm h})$ to just above 1.25. The third anomalous galaxy enters the primary halo and obtains a peaked velocity profile during a starburst associated with its first pericentric passage. In the remaining two anomalous galaxies we find that a bar drives the growth of a central stellar mass excess at high redshift ($z > 5$). At those times the bar is not properly resolved and, hence, the corresponding increase in V_c may be, at least in part, of numerical origin. However, bars form also at later times (Figure 2.2 clearly shows that many galaxies at $z \sim 3$ are barred) and can drive the steady increase of the peakedness of the circular velocity profile.

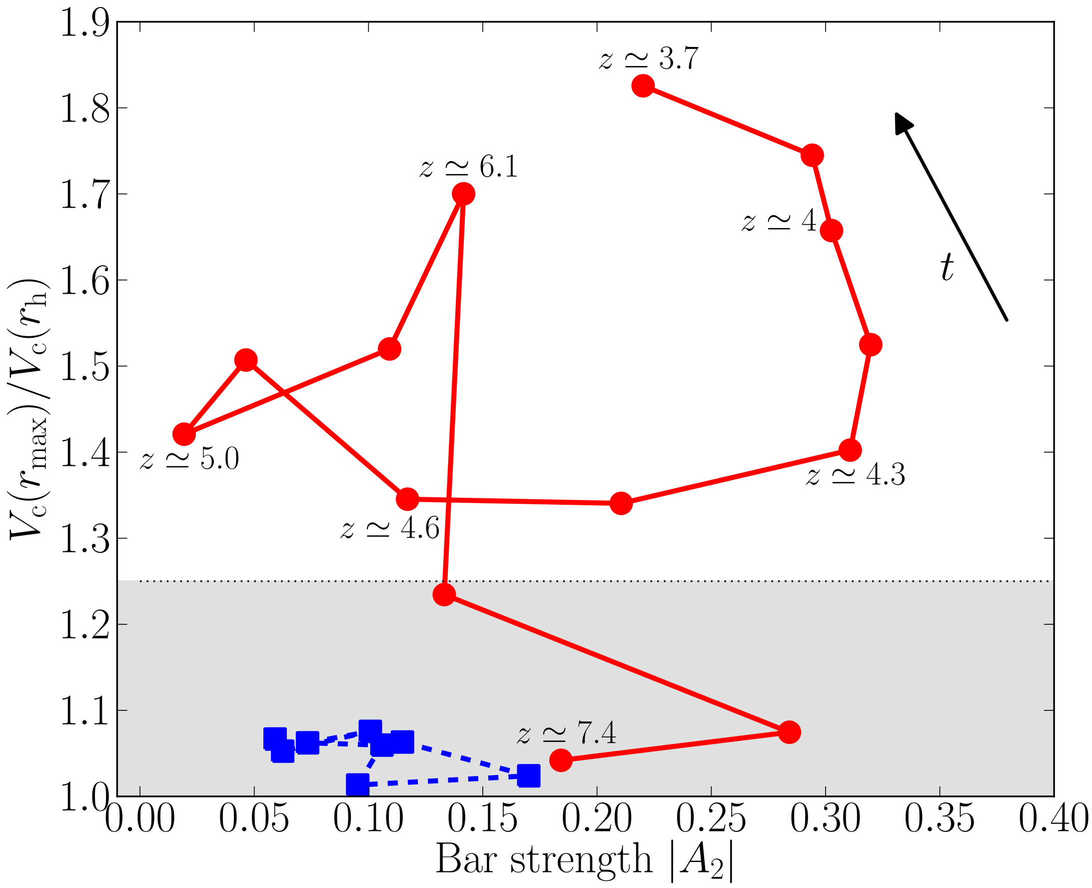

Figure 2.11 Evolution of $V_c(r_{max})/V_c(r_h)$, the peakedness of the circular velocity profile, as a function of the bar strength $|A_2|$. The red solid curve corresponds to the third galaxy in the first row of Figure 2.4. This galaxy has a peaked rotation curve at $z \sim 3.7$, but the growth of $V_c(r_{max})/V_c(r_h)$ is not associated with a major merger event. Instead, it gets peaked as a response to the formation of a stellar bar. Reference redshifts are marked in the figure and the arrow indicates the direction of time along the curve. The blue dashed curve shows the corresponding result for the fourth galaxy in the third row. This is a "control" galaxy with a flat circular velocity curve. As expected no strong bar develops in this case.

As an example of the this process, we plot in Figure 2.11 the evolution of the bar strength for one of the galaxies (third galaxy in the first row of Figure 2.2). The bar strength is defined following Dubinski, Berentzen & Shlosman (2009) as the absolute value:

$$|A_2| = \frac{1}{M}\left|\sum_{j=1}^{N} m_j \exp(2i\phi_j)\right|, \tag{2.3}$$

where the summation is performed over the N star particles within a cut-off radius $R_\mathrm{c} =$ 500 pc from the centre of the galaxy, m_j and ϕ_j are, respectively, the mass and the polar angle in the plane of the galaxy of the j-th star particle, and $M = \sum_{j=1}^{N} m_j$. The figure shows the evolution of $V_\mathrm{c}(r_\mathrm{max})/V_\mathrm{c}(r_\mathrm{h})$, the peakedness of the circular velocity profile, as a function of the bar strength $|A_2|$. We find that the peakedness increases *after* the bar forms, i.e., as a response to the presence of the bar. In turn, the bar gets weaker as the peakedness increases – possibly because the formation of a bulge shuts down the passage of waves through the central region via the formation of an inner Lindblad resonance which reduces the effectiveness of repeated swing amplification (e.g., Toomre 1981; Sellwood 1989).

The discussion above is summarised by Figure 2.12, which shows the distribution of $V_\mathrm{c}(r_\mathrm{max})/V_\mathrm{c}(r_\mathrm{h})$ for our galaxy sample. Figure 2.12 suggests that major mergers might be the most effective process in shaping (and steepening) the central part of the circular velocity curve. Given the connection between V_c and morphology discussed in Section 2.3.4, this in turn suggests that major mergers may be the primary process in assembling bulges with $B/T \geq 0.3$, in agreement with previous theoretical works (e.g. Naab & Trujillo 2006; Hopkins et al. 2010; Kraljic, Bournaud & Martig 2012). This is valid at least at high redshift ($z \geq 3$), in slightly over-dense, group/proto-cluster environments and at mass scales ($M_\star \lesssim 10^{10}$ $\mathrm{M}_\odot$) below the exponential cutoff mass of the stellar mass function at $z > 2$ (Pérez-González et al., 2008; Marchesini et al., 2009; Muzzin et al., 2013; Tomczak et al., 2014).

2.4 Caveats

We probe the morphological evolution of a population of high-redshift galaxies formed in the Argo zoom-in cosmological simulation. A key feature of this simulation is the high resolution and the big zoom-in sub-volume that allows us to follow the assembly of a normal galaxies population from redshift ≤ 10 down to $z \simeq 3$. However, our simulation does not seem to be able to reproduce all the properties of the observed high redshift galaxies, as we discuss in the following of this section. Nonetheless, we argue that our results are, at least in the qualitative picture, independent of and only mildly afflicted by these shortcomings.

We showed that the shape of the circular velocity curve correlates with the B/T ratio and with stellar mass. McCarthy et al. (2012) found similar results in terms of circular velocity curves' shape and trends with mass, but they attributed them to spurious overcooling effects above a certain stellar mass. Although we can not exclude some residual

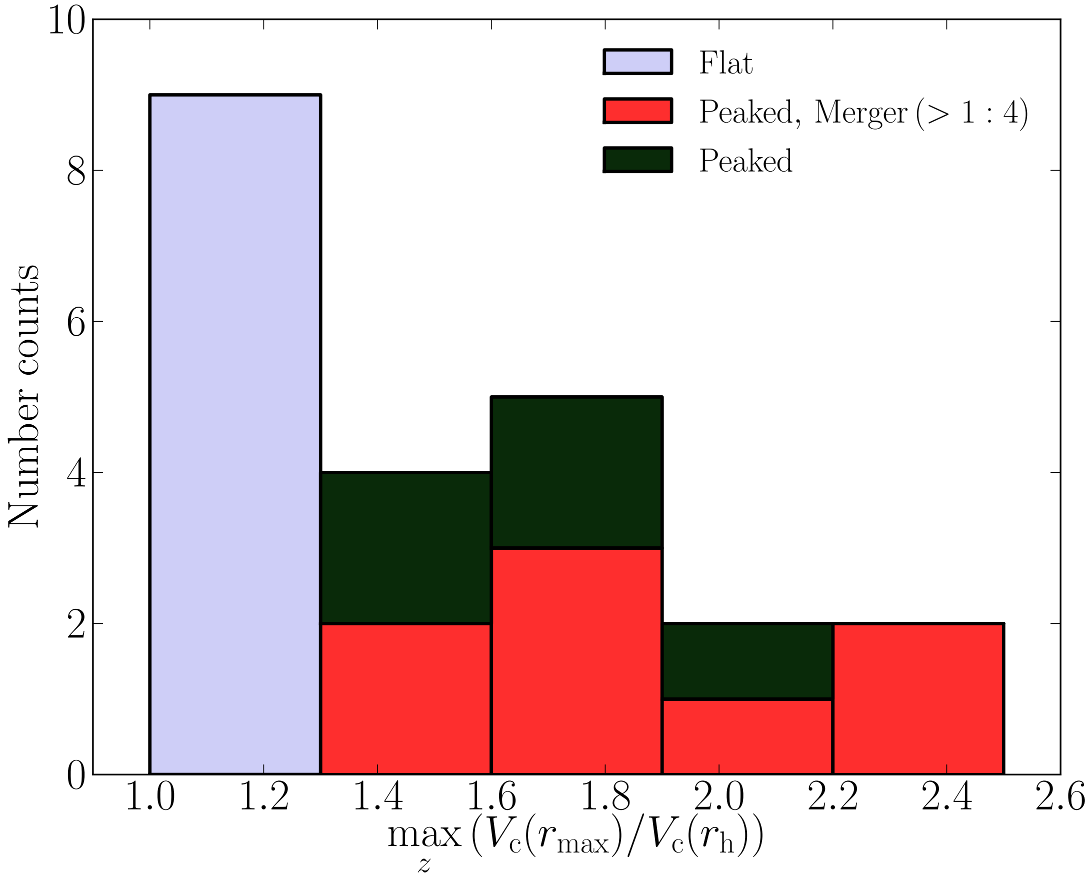

Figure 2.12 Histograms of the distributions of the maximum $V_c(r_{\max})/V_c(r_h)$ over z. Light blue bars show the number of galaxies with always flat rotation curves (9 out of 22). Red bars indicate the distribution of galaxies that become peaked as a result of a major merger (8 out of 22). Dark green bars show the distribution of galaxies that become peaked for other reasons (5 out of 22), see text. Most galaxies that have a peaked circular velocity curve at late times undergo a major merger. In contrast, none of the galaxies with a flat circular velocity profile undergo a major merger with mass ratio $\geq 1:4$.

overcooling, we are confident that this has only a minor impact on our results, because (i) Argo has a much higher resolution with respect to the simulation presented in McCarthy et al. (2012), in particular our softening is always smaller than the effective radius of the galaxies that we analyse, (ii) the most massive galaxy in our simulation, in which artificial overcooling should be most severe, matches nicely the predictions of abundance matching (see Figure 2.13 below), (iii) we can attribute, in most cases, a physical cause to the sudden transformations of our galaxies, in particular to the increase of the central density and the associated steepening of the circular velocity curve.

A more serious challenge for our simulation is to reproduce the $M_\star - M_h$ relation that is empirically determined with the help of the abundance matching technique (Behroozi, Wechsler & Conroy, 2013; Moster, Naab & White, 2013). In Figure 2.13 we show the

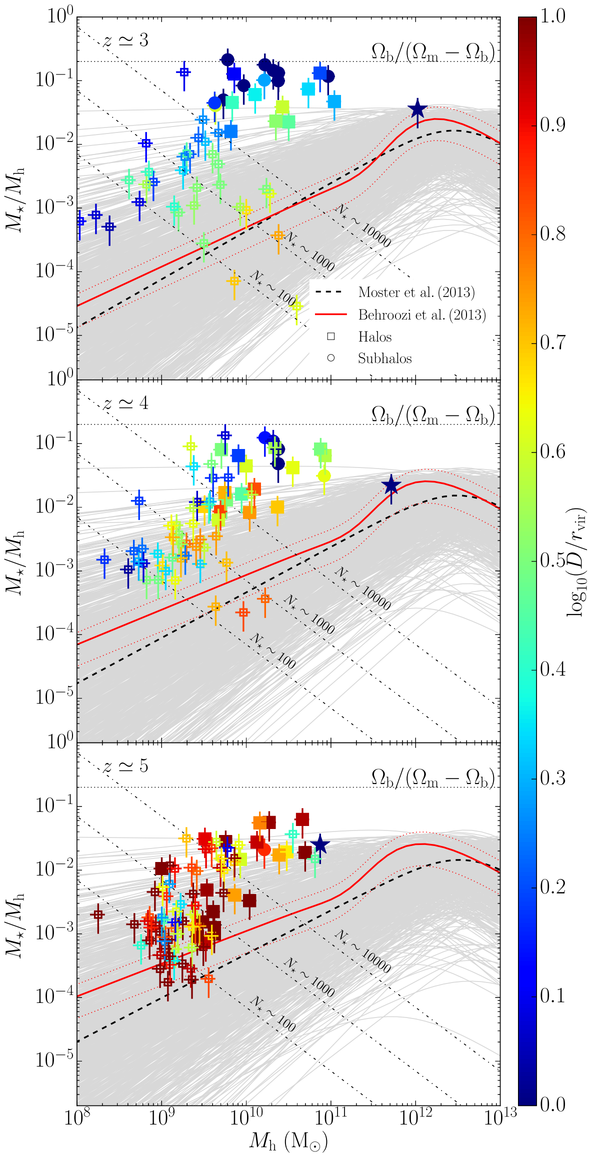

Figure 2.13 Stellar mass – halo mass relation at $z \simeq 3$ (top), $z \simeq 4$ (middle) and $z \simeq 5$ (bottom). Large filled symbols show the galaxies in our sample. Circles mark satellite galaxies and squares central galaxies. Small empty symbols refer to additional central galaxies at each redshift with at least 100 star particles. The star marks the central galaxy of the primary (group-sized) halo. Symbols are colour coded according to the distance of each galaxy from the primary halo in units of the virial radius of the primary. Error bars assume 50% and 20% uncertainty on the x and y axis, respectively (Munshi et al., 2013). The red continuous and the black thick dashed lines show the abundance matching from Moster, Naab & White (2013) and Behroozi, Wechsler & Conroy (2013), respectively. Dash-dotted lines show three different resolution limits for a given, approximate number of star particles.

$M_\star - M_h$ relation at $z \simeq 3$, 4 and 5 for the galaxies of our sample and for all central galaxies in the high resolution region of the Argo simulation that contain at least 100 star particles. Stellar masses are measured within 0.1 $r_{\rm vir}$ excluding any contributions from satellites. M_h is typically the virial mass of the parent dark matter halo hosting the galaxy. However, for the satellite galaxies in our sample we use the halo masses at infall into their first host halo.

The figure shows that a large number (but not all) of the simulated galaxies lie above the empirically determined $M_\star - M_\mathrm{h}$ relation. This discrepancy (i) grows with time, (ii) is larger for galaxies closer to the primary (group-sized) halo and (iii) is more pronounced for satellite galaxies than for central galaxies. We therefore suggest that we are witnessing to some extent an environmental effect. In particular, there are hints that the dark matter halo of galaxies in the vicinity of the primary halo might have undergone suppressed growth relative to the halo of a central galaxy of the same stellar mass, an effect ultimately exacerbated by progressive tidal truncation as galaxies enter the virial radius. Further evidence supporting this interpretation is that the distance dependence of the discrepancy grows with time (see Figure 2.13). If insufficient feedback were the single cause of the mismatch we would not expect to find a dependence on distance from the primary halo nor any time evolution of such a correlation. Likewise, it is conceivable that there is an interplay between such effects and feedback, and therefore that the magnitude of such effects and thus the stellar mass assembled at a given time is still somewhat affected by the strength and mode of the feedback adopted. Finally, as we mentioned above, the central galaxy, for which insufficient feedback should be especially problematic, nicely matches abundance matching predictions (see also Feldmann & Mayer 2014). We plan to study the effect of the environment on the $M_\star - M_\mathrm{h}$ relation in a forthcoming paper.

However, many of more massive galaxies at $z \simeq 5$ tend to be inconsistent with the abundance matching predictions regardless of the distance from the main halo. It is also conceivable to expect that adding metal-line cooling would exacerbate the problem by increasing further the star formation efficiency, as suggested, for example, by the somewhat higher mass of the spiral galaxy in the ErisMC simulation (Shen et al., 2012), which included metal-line cooling, relative to the original Eris with only primordial cooling (Shen et al., 2012). While these galaxies have a somewhat biased formation history, given that they start within a slightly overdense Lagrangian region of the Universe and end up close to the primary halo at $z = 3$, it is nonetheless plausible that the feedback model in our simulations is not efficient enough in low mass galaxies. A conclusion along these lines has been suggested by Stinson et al. (2013), who argue that early stellar feedback (e.g., radiation pressure from massive stars) needs to be included in order to reduce the stellar fraction of high redshift galaxies. The need for a different mode of feedback that delays star formation might be indicated also by the gas fractions found in the simulated galaxies. These are in the range $\sim 5 - 70\%$. In a number of cases, they fall below the typical values found for high-z galaxies, which are in the range $20-80\%$ of the stellar mass (see e.g. Daddi et al. 2010; Tacconi et al. 2010; De Breuck et al. 2014). Nevertheless, we caution that measurements of gas fractions are usually available for galaxies significantly brighter than those in our sample, and they span a somewhat lower redshift, tipically below $z \sim 2.5$ (Daddi et al., 2010; Tacconi et al., 2010; De Breuck et al., 2014).

Despite the aforementioned shortcomings, we maintain that the physical origin (major mergers) of the identified morphological transformations implies that our findings are robust to changes of the feedback physics. Nonetheless, it is likely that changes to the feedback model affect the timescales over which the morphologies of galaxies settle into their $z = 0$ state. It will clearly be helpful to compare our results with those of recent cosmological simulations that aim at modelling a variety of feedback processes at high resolution (e.g., Hopkins et al. 2013b).

2.5 Discussion and conclusions

We investigated the properties of a sample of 22 high-z galaxies during the assembly of a galaxy group in the Argo zoom-in cosmological simulation. We focused on the morphological evolution of this galaxy sample as measured by a variety of (correlated) diagnostics such as the stellar surface-density profile, the B/T ratio, the shape of the circular velocity profile and a visual morphological classification. We found that the peakedness of the circular velocity profiles is a good proxy for the morphological classification. In particular, galaxies with “peaked” circular velocity curves correspond to systems with more massive bulge components, while “flat” curves correspond to bulge-less disc galaxies of typically low stellar masses. At the intermediate mass scale ($M_\star \lesssim 10^9$ M$_\odot$ and $M_{\rm vir} \sim 0.5-5\times10^{10}$ M$_\odot$) a wide variety of rotation curves and B/T ratios are present, showing that (halo) mass alone does not completely determine morphology. Instead, the stellar mass is more closely correlated with morphology.

By analysing the origin of the dichotomy between “flat” and “peaked” galaxies we have identified major mergers as the main evolutionary trigger, with disc instabilities and minor accretion/interactions of satellites playing a sub-dominant role. Note that this conclusion echoes the finding of Feldmann, Carollo & Mayer (2011) for the evolution of a galaxy population at much lower redshift ($z \sim 0 - 1.5$); indeed they identified mergers prior to infall inside the group potential as the main culprit behind the transmutation from discs into spheroids.

Our main conclusion is thus two-fold; (i) galaxy evolution in typical environments at high-z appears to be surprisingly similar to galaxy evolution at low-z despite the fact that galaxies are inherently more gas-rich and are fed by the cosmic web at higher gas accretion rates, and (ii) the Hubble Sequence is established very early during galaxy assembly, at $z \sim 3 - 4$, at least in the slightly biased regions around galaxy groups.

The central mass excess in our galaxies with peaked circular velocity profiles often appear to be exponential pseudobulges based on their stellar density profiles (although stellar bars can also contribute). One notable exception is the central galaxy, which exhibits a profile that is consistent with a de Vaucouleurs profile as expected for massive early-types, probably as a result of several repeated mergers/interactions (Naab & Trujillo, 2006; Cox et al., 2006; Moster et al., 2011). This suggests that a mixture of dynamical and secular processes might be responsible for the formation of pseudobulges, supporting the findings of Guedes et al. (2013).

In the galaxies of our sample we do not observe the formation of giant star forming clumps via violent disc instabilities (Noguchi, 1999; Dekel, Sari & Ceverino, 2009; Ceverino, Dekel & Bournaud, 2010). However, all but one of the galaxies in our sample are of low and intermediate mass, while clumpy galaxies typically belong to the massive galaxy population. Furthermore, our simulations do not allow gas cooling below $\sim 10^4$ K and the higher pressure hinders the gravitational collapse via a local Toomre instability (Ceverino, Dekel & Bournaud, 2010). In addition, the effective stellar feedback in our simulation potentially suppresses the life times of giant clumps (Hopkins et al., 2012, 2013a; Moody et al., 2014). While galaxies with massive clumps are thus probably not the typical galaxies at high-z, a clumpy phase may still play an important role in the evolution of massive central galaxies in high density environments, e.g., for progenitors of present-day central group and cluster galaxies.

Our results suggest that the morphology of high redshift galaxies is determined by processes similar to those operating on/in local galaxies. In particular major mergers, a natural consequence of the hierarchical structure formation in a ΛCDM Universe, appear to play a crucial role in the early morphological transformation of galaxies. Future observations with ALMA, E-ELT, or JWST as well as larger samples of properly resolved, simulated galaxies may shed more light on the interplay between merging, feedback, and galaxy morphology as function of stellar mass and environment in high redshift galaxies. Surveys with these new ground based and space born instruments will open new observational windows into the typical galaxy population at high redshift and be able to test our prediction that the Hubble Sequence near groups and clusters is already in place by $z = 2$.

3

Growing massive black hole seeds within quasi-stars: the impact of outflows and rotation[1]

Abstract In this Chapter I will discuss the structure and evolution of quasi-stars, i.e. $\sim 10^6$ $\mathrm{M}_\odot$ gaseous envelopes sustained against their own self gravity by the accretion luminosity produced by a central ~ 100 $\mathrm{M}_\odot$ black hole. I will develop an analytical model to describe super-Eddington continuum-driven winds and I will adapt it to quasi-stars. I will show that (i) the outflows can disperse the envelope mass faster than what the black hole can accrete, and (ii) the winds are not "photon-tired" and dim, i.e. quasi-stars are expected to be bright objects, possibly detectable with James Webb Space Telescope. Then, I will include rotation in the model; specifically, I will explore the conditions to form a central accretion disc. Such conditions are often not attained within quasi-stars; therefore, I will speculate on the evolutionary transition between supermassive stars and quasi-stars and on the formation of massive black hole seeds when the quasi-star phase is absent.

---

[1] The material of this chapter is contained in and adapted from the following publications: **Fiacconi** & Rossi 2016, MNRAS, 455, 2; **Fiacconi** & Rossi, submitted to MNRAS

3.1 Introduction

Supermassive black hole formation is an outstanding question in astrophysics (see also Section 1.2.4). The compelling evidence that links supermassive black holes' evolution to that of their host galaxies (e.g. Magorrian et al. 1998; Ferrarese & Merritt 2000; Tremaine et al. 2002; Marconi & Hunt 2003; Gültekin et al. 2009; McConnell & Ma 2013), strongly suggests that the answer must be sought in the broader context of galaxy assembly.

During the last ten years or so, observations have unambiguously proved the existence of supermassive black holes accreting at the centre of bright quasars at redshifts $z \gtrsim 6$ with masses in excess of 10^9 $\mathrm{M}_\odot$ (Fan et al., 2006; Willott et al., 2010; Mortlock et al., 2011; Wu et al., 2015). Despite that those objects are not perhaps representative of the entire population of supermassive black holes at $z \gtrsim 6$ (e.g. Treister et al. 2013; Weigel et al. 2015), they represent a challenge for many theoretical models that attempt to describe the formation of the first black hole seeds. Indeed, black hole seeds originating both as the leftovers of the first population (PopIII) stars (with masses $\lesssim 100$ $\mathrm{M}_\odot$; Madau & Rees 2001; Tanaka & Haiman 2009), and as the product of dynamical processes at the centre of primordial nuclear star cluster (with masses $\lesssim 1000$ $\mathrm{M}_\odot$; Quinlan & Shapiro 1990; Devecchi & Volonteri 2009; Devecchi et al. 2012), are not expected to grow fast enough to reach $\sim 10^9$ $\mathrm{M}_\odot$ by $z \sim 6$, unless they experience prolonged periods of super-Eddington accretion (e.g. Madau, Haardt & Dotti 2014; Volonteri, Silk & Dubus 2015).

Although these observational constraints do not necessarily rule out those mechanisms on a physical base, high-z quasars may more easily be explained by the so called "direct collapse" scenario (e.g. Bromm & Loeb 2003; Begelman, Volonteri & Rees 2006; Lodato & Natarajan 2006; Dijkstra et al. 2008; Begelman & Shlosman 2009; Latif et al. 2013; Mayer et al. 2015). This latter envisages a large mass of pristine gas ($\sim 10^6 - 10^7$ $\mathrm{M}_\odot$), promptly accumulated at the centre of a galaxy-size halo on (sub)parsec scales. A large fraction of it would rapidly ($< 10^6$ yr) form a massive seed ($10^4 - 10^6$ $\mathrm{M}_\odot$), directly at the centre of a galaxy at $z \sim 15$. Attractive as it is, this process is far from being proved and at least two major steps require further investigations.

Although in principle there is plenty of gas available at high redshift and cold flows have been shown to be effective in bringing that gas down to the centre of (massive) haloes (Di Matteo et al., 2012), the conditions to avoid substantial fragmentations and to overcome the centrifugal barrier are not fully understood yet. Several possibilities have been discussed, such as the dissociation of H_2 molecules by Lyman-Werner ionising radiation coming from nearby, star-forming galaxies in order to avoid cooling and fragmentation (Ferrara & Loeb, 2013; Dijkstra, Ferrara & Mesinger, 2014). Another possibility is the onset of supersonic turbulence and the removal of angular momentum due to non-axisymmetric perturbations and gravitational torques during the collapse of the halo (Begelman & Shlosman, 2009; Choi, Shlosman & Begelman, 2013, 2015), or at the centre of major merger remnants between rare and massive galaxies at high redshift (Mayer et al., 2010, 2015).

The second issue is how to actually form a black hole and what is its initial mass. The answer may vary according to the physical properties of the assembled mass. When more than $\sim 10^8$ $\mathrm{M}_\odot$ can be rapidly piled up, the resulting structure likely becomes dynamically unstable (even if rotating) and relativistic radial instability can lead to implosion and direct black hole formation (Fowler, 1966; Baumgarte & Shapiro, 1999; Shibata & Shapiro, 2002). However, forming such a structure requires rather extreme conditions (e.g. a major merger; Mayer et al. 2015). With relatively milder accretion rates, nuclear burning can start at the centre of a convectively stable object, i.e a supermassive star (Begelman, 2010). After a million years, the core that could not convectively acquire fresh hydrogen collapses to form a stellar size (~ 100 $\mathrm{M}_\odot$) black hole. Highly optically thick gas keeps however falling onto the newly born black hole, possibly with enough angular momentum to be able to generate accretion power. This energy feedback inflates the innermost part of this inflow, creating a *quasi-star*: a massive, slowly rotating envelope, sustained against its own gravity by the black hole accretion power (Begelman, Rossi & Armitage, 2008; Begelman, 2010; Volonteri, 2010; Ball et al., 2011; Dotan, Rossi & Shaviv, 2011). At quasi-star centres, the embryo black holes may accrete at a super-Eddington rate (at about the Eddington rate for *the envelope mass*, which is typically much larger than the black hole mass), as energy is transported outward by convection (not by radiative diffusion) through the envelope. The broad expectation is that seeds of $\sim 10^4 - 10^5$ $\mathrm{M}_\odot$ may easily grow in $\lesssim 1$ Myr (Begelman, Rossi & Armitage, 2008). After this time, the envelope would be definitively dispersed and accretion would proceed at an Eddington limited fashion directly from the protogalactic disc.

Content of the Chapter In the following, we study the structure and the evolution of quasi-stars; specifically we explore the role of outflows and rotation. First, we focus on super-Eddington, continuum-driven winds in Section 3.2 and in Section 3.3 we apply them to describe the hydrostatic structure of quasi-stars and to assess the impact of outflows on their evolution. We find that outflows may represent a bottleneck for the growth of the central black holes. Then, in Section 3.4 we explore the effects of rotation on the stability of quasi-stars, specifically by looking at the conditions to form a central accretion disc, a necessary feature for the self-consistency of the model. We find that it is often not possible to form an accretion disc and we then speculate in Section 3.4.4 on the fates of supermassive stars at the verge of the collapse of their central core, i.e. just before that a quasi-star would be expected to form. In section 3.5, we summarise our findings and conclude.

3.2 The wind model

3.2.1 Equations and general properties

We consider a stationary, spherically-symmetric, radiation-dominated wind launched by a non-rotating, stellar-like object of mass $M_\star$ from a spherical surface of radius $R_\star$, that represents the base of the wind[2]. By radiation-dominated, we mean that the contribution of the gas pressure $p_{\rm gas}$ is assumed to be negligible compared to the radiation pressure $p_{\rm rad}$, i.e. $p_{\rm gas}/p_{\rm rad} \ll 1$. This assumption allows us to neglect the presence of $p_{\rm gas}$ in the following calculations (and we simply write $p \equiv p_{\rm rad}$), but it requires at the same time that the wind is launched from a radiation-dominated object. Just outside $R_\star$, the gas is assumed to be initially optically thick and interacts with radiation through a (constant) opacity κ. We are interested in primordial composition objects, where line-driven interaction is negligible. The equations that describe this system are similar to those used by several previous works about stellar winds (and spherical accretion) in both the optically thin and optically thick regime (e.g. Żytkow 1972; Begelman 1978, 1979; Kato 1983; Quinn & Paczynski 1985):

$$\dot{M} = 4\pi r^2 \rho\, v, \tag{3.1}$$

$$\frac{1}{2}\frac{\mathrm{d}v^2}{\mathrm{d}r} = -\frac{GM_\star}{r^2} + \frac{\kappa L}{4\pi r^2 c}, \tag{3.2}$$

$$\dot{M}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{v^2}{2} - \frac{GM_\star}{r} + \frac{p+U}{\rho}\right) = -\frac{\mathrm{d}L}{\mathrm{d}r}. \tag{3.3}$$

These equations determine the structure of the gas density ρ, the radial gas velocity v, the luminosity carried by photons L, the (radiation) pressure p and the (radiation) internal energy density U as a function of the spherical radius r within the gravitation potential $\Phi = -GM_\star/r$ induced by $M_\star$ outside $R_\star$. The steady-state wind is characterised by the constant outflow rate $\dot{M}$. Equations (3.1), (3.2) and (3.3) describe the conservation of mass, momentum, and energy, respectively.

Such a system of equations is not closed and several approaches can be used to close it to different degrees of approximation. Shaviv (2001b) and Owocki, Gayley & Shaviv (2004) start from similar equations, except that they initially include the contribution of the gas thermal pressure to the momentum and energy conservation. Then, they simplify the system focusing on the supersonic branch, thus subsequently neglecting the gas pressure terms in the momentum equation (which brings it back to our same equation 3.2) and the advection term $(p+U)/\rho$ (with the corresponding one due to gas pressure) in the energy equation. This approximation leads to the great advantage that fully analytic solutions can be derived. However, the limitation is that the lack of the advective term makes the behaviour of the wind insensitive to the local optical thickness.

---

[2] In the following, we will always use the subscript $_\star$ to indicate quantities evaluated at $R_\star$

Instead, we follow an approach similar to that used by e.g. Quinn & Paczynski (1985) and we explicitly include additional prescriptions to properly describe the behaviour of the wind in the extrema of very optically thin and optically thick regime, i.e. when the optical depth:

$$\tau(r) = \int_{r}^{+\infty} \kappa \; \rho(x) \; \mathrm{d}x, \tag{3.4}$$

is either $\tau \ll 1$ or $\tau \gg 1$, respectively. When the wind is optically thick, radiation and matter can reach local thermodynamical equilibrium at the same temperature T (which relates to the energy density $U = 3p = aT^4$, where a is the radiation constant) and the gradient of the radiation energy density is:

$$\left.\frac{\mathrm{d}U}{\mathrm{d}r}\right|_{\tau\gg1} = -\frac{3\kappa\rho L}{4\pi r^2 c}. \tag{3.5}$$

On the other hand, local thermodynamical equilibrium may not be reached in the optically thin limit and a unique temperature may not be a physically-motivated quantity. In this case, the photons carrying L travel with roughly radial orbits and interact very little with matter, keeping L almost constant (see e.g. Żytkow 1972). Then, the radiation energy density decreases mostly because of geometrical dilution in a progressively larger volume:

$$\left.\frac{\mathrm{d}U}{\mathrm{d}r}\right|_{\tau\ll1} = -\frac{L}{2\pi r^3 c}. \tag{3.6}$$

We follow Quinn & Paczynski (1985) in defining the total gradient of U as the sum of the two limiting cases:

$$\frac{\mathrm{d}U}{\mathrm{d}r} = -\frac{L}{2\pi r^3 c} \; f(\tilde{\tau}), \tag{3.7}$$

where we define the function:

$$f(\tilde{\tau}) = \frac{3\tilde{\tau}}{2} + 1, \quad \tilde{\tau} \equiv \kappa \; \rho \; r. \tag{3.8}$$

The "effective" opacity $\tilde{\tau}$ leads the gradient of U to the right optically thin and optically thick limits when $\tilde{\tau} \ll 1$ and $\tilde{\tau} \gg 1$, respectively. However, $\tilde{\tau}$ is just an approximation of the actual opacity τ; the two are related by a constant factor when τ is a power law and such a factor is close to 1 when $\tau \propto r^{-1}$. Although we do not know a priory the relationship between τ and $\tilde{\tau}$, we demonstrate in the following that $\tau \propto r^{-1}$ roughly holds and therefore $\tilde{\tau} \simeq \tau$ (see Section 3.2.3 and 3.3).

Finally, we need to relate p and U to close the system of equations. Quinn & Paczynski (1985) implicitly assume that $p = U/3$ everywhere in the flow (see their equations 11b and 12). This is correct in the optically thick regime, but is not valid when the gas is optically thin. Indeed, $U = p$ when the gas is optically thin; this different relation between p and U is also responsible for the inexact relation between the luminosity observed by an observer at infinity and by an observer comoving with the flow, as reported by Quinn & Paczynski (1985; see also Section 3.2.2 and Cassinelli & Castor 1973). In order to have a smooth transition between the two regimes, similar to the case of the gradient of U (see equation 3.7), we propose the following functional form for the opacity-dependent ratio p/U:

$$\frac{p}{U} \equiv g(\tilde{\tau}) \equiv \left(\frac{3\tilde{\tau}}{2}+1\right)\left(\frac{9\tilde{\tau}}{2}+1\right)^{-1}. \tag{3.9}$$

Figure 3.1 shows the behaviour of both $f(\tilde{\tau})$ and $g(\tilde{\tau})$. The latter goes correctly from 1 (when the wind is optically thin) to 1/3 (when the wind is optically thick). However, the accuracy of both prescriptions is questionable around $\tilde{\tau} \sim 1$, because the actual form is largely arbitrary. We therefore compute and compare wind models, choosing different functional form for $f(\tilde{\tau})$ and $g(\tilde{\tau})$ and concluded that our results are not affected as long as the the limits are correct and the transition occurs rapidly (over several $\tilde{\tau}$) around $\tilde{\tau} \sim 1$.

The equation of the conservation of energy can be directly integrated, becoming an algebraic equation for L:

$$L(r) = \dot{E} - \dot{M}\left(\frac{v^2}{2} - \frac{GM_\star}{r} + (1+g(\tilde{\tau}))\frac{U}{\rho}\right), \tag{3.10}$$

where we make use of $g(\tilde{\tau})$ explicitly and we introduce the total conserved luminosity $\dot{E}$, which represents the constant of integration. The system of equations that we finally solve (often dubbed as "wind equations" in the following) is composed of equations (3.1), (3.2), (3.7) and (3.10), coupled with the definitions of $\tilde{\tau}$, $f(\tilde{\tau})$ and $g(\tilde{\tau})$. The system has two ordinary differential equations and two algebraic equations for the dependent variables ρ (or $\tilde{\tau}$), v, L and U as a function of r.

At this point, it is convenient to introduce new dimensionless variables. We define the new velocity variable $w = v^2/v_{\rm esc}^2$, where $v_{\rm esc}^2 = 2GM_\star/R_\star$ is the escape velocity from the base of the wind, the new radiation energy density variable $u = U\kappa R_\star^2/(GM_\star)$, and the new luminosity variable $\Gamma = L/L_{\rm Edd}$. Γ is the Eddington ratio and the Eddington luminosity $L_{\rm Edd}$ is defined as:

$$L_{\rm Edd} = \frac{4\pi c G M_\star}{\kappa} = 1.26\times 10^{38}\ \tilde{\kappa}^{-1}\ m_\star\ \mathrm{erg\ s^{-1}}, \tag{3.11}$$

where $\tilde{\kappa}$ is the opacity in units of the electron scattering opacity $\kappa_{\rm es} = 0.35\ \mathrm{cm^2\ g^{-1}}$ (assuming primordial abundances) and $m_\star$ is the stellar mass $M_\star$ in units of solar masses. The independent variable r can also be transformed into $x = 1 - R_\star/r$, such that the interval $r \in [R_\star, +\infty)$ is mapped into $x \in [0,1)$. We can first express equation (3.2) with the new variables as:

$$w' = \Gamma - 1, \tag{3.12}$$

where here and in the following $' = \mathrm{d}/\mathrm{d}x$. The gradient of the radiation energy density u becomes:

$$u' = -2\ (1-x)\ \Gamma\ f(\tilde{\tau}), \tag{3.13}$$

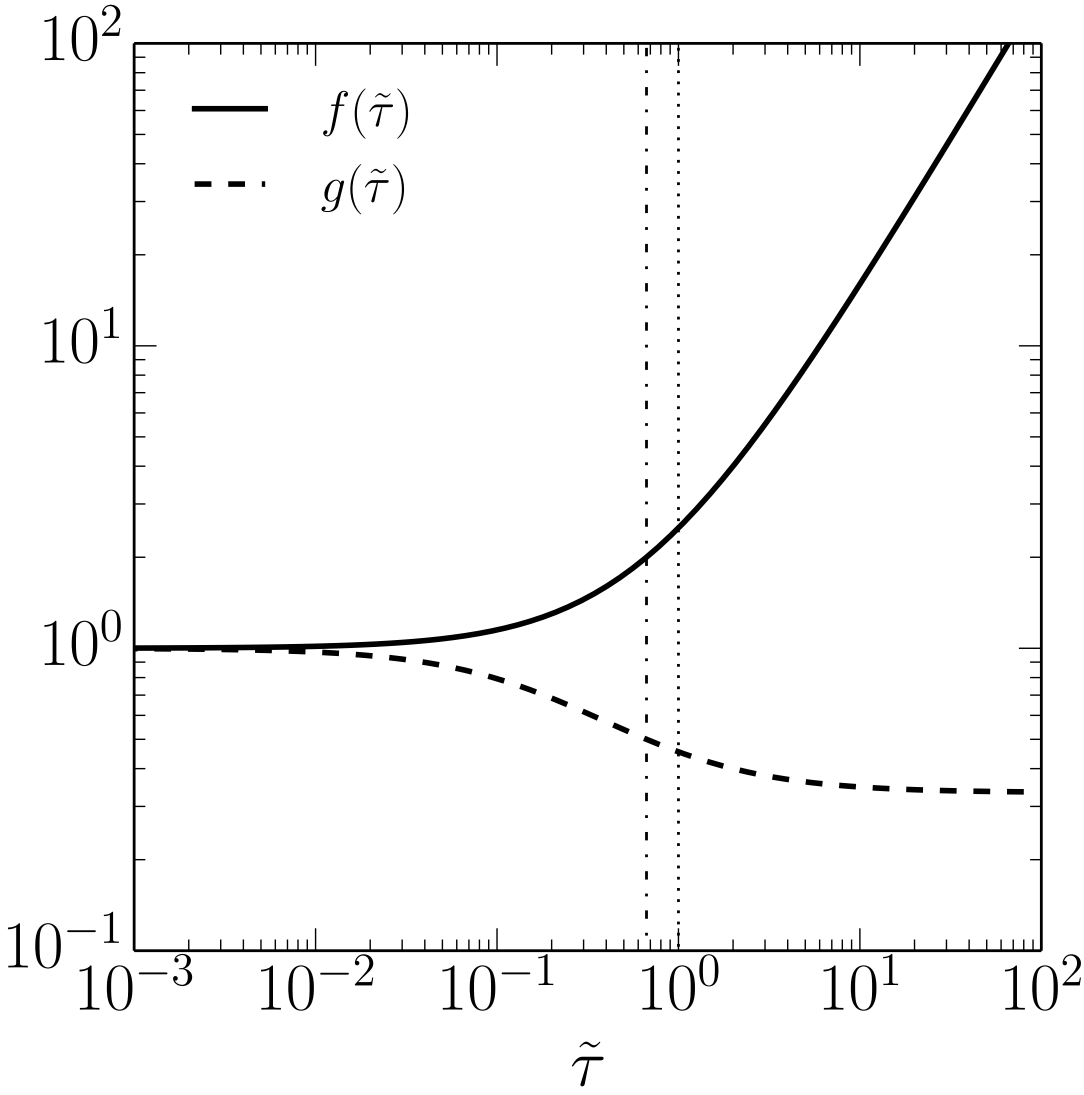

Figure 3.1 The tick continuous and dashed lines show the behaviour of f and g as a function of $\tilde{\tau}$, respectively. For reference, the vertical dotted and dashed lines mark the points $\tilde{\tau} = 1$ and $\tilde{\tau} = 2/3$, respectively.

where $f(\tilde{\tau})$ is defined in equation (3.8) and:

$$\tilde{\tau} = \frac{\alpha\beta}{w^{1/2}}(1 - x), \tag{3.14}$$

is the definition of $\tilde{\tau}$ using our dimensionless variables. We introduce the two factors α and β; α is a dimensionless expression for $\dot{M}$ in terms of the trapping radius $R_{\rm tr}$ (Begelman, 1978, 1979):

$$\alpha = \frac{R_{\rm tr}}{R_\star} = \frac{\kappa \dot{M}}{4\pi c R_\star}. \tag{3.15}$$

The trapping radius is where the diffusion time scale for photons is equal to the dynamical time of the outflowing wind, $R_{\rm tr}(\tilde{\tau}/c) \approx R_{\rm tr}/v$, which implies that $\tilde{\tau} \approx c/v$ at $R_{\rm tr}$. We will see in the following that within this radius, since the radiation is *trapped*, the luminosity transported by diffusion becomes subdominant with respect to the energy advected within the flow. The parameter β is a dimensionless factor depending on the properties of the star only:

$$\beta \equiv \frac{c}{v_{\rm esc}} \approx 486\ m_\star^{-1/2}\ r_\star^{1/2}, \tag{3.16}$$

where $r_\star$ is the stellar radius in units of solar radii. β appears naturally from the normalisation of the optical depth and measures the deepness of the gravitational potential well of the parent star. It depends on the stellar properties $R_\star$ and $M_\star$ only and therefore, $M_\star$ and β (or $v_{\rm esc}$) are enough to rescale the equations in physical units. Finally, the algebraic equation for Γ can be obtained from equation 3.10 and reads:

$$\Gamma = \dot{\mathcal{E}} - \alpha \left(w + x - 1 + \frac{1 + g(\tilde{\tau})}{1 - x} \frac{u}{\tilde{\tilde{\tau}}} \right), \tag{3.17}$$

where $\dot{\mathcal{E}} = \dot{E}/L_{\rm Edd}$ and $g(\tilde{\tau})$ is defined in equation (3.9). The dependent variables w, u and Γ are proportional to the kinetic energy of the gas, to the energy density of the radiation and to the luminosity carried by photons, respectively. Therefore, equations (3.12), (3.13) and (3.17) compose the system that describes the energy exchanges between the different components of the system.

3.2.2 Numerical integration of the wind equations

We integrate numerically the wind equations (equations 3.12, 3.13, 3.14 and 3.17) using the CVODE module of the SUNDIALS[3] package (Cohen, Hindmarsh & Dubois, 1996; Hindmarsh et al., 2005). CVODE is a C solver for stiff and non-stiff ordinary differential equation systems in explicit form. We adopt a fifth-order backward differentiation formula in fixed-leading coefficient form with a modified Newton iteration to solve non-linear systems. CVODE provides also a module to find the roots of nonlinear equations which is well suited to determine the position of the photosphere and the local properties of the wind while contemporary solving the wind equations.

In order to find solutions of the wind equations, we follow the procedure outlined by Quinn & Paczynski (1985; see also Żytkow 1972; Kato 1983). First of all, we characterise our star by choosing a value for β. Then, we pick a value for α and $\dot{\mathcal{E}}$. These two constants of integration are not enough to fully characterise the wind. We need a boundary condition, specifically the asymptotic gas velocity at infinity w_∞. With that, we can specify the initial conditions at a large radii $x_\infty = 1 - \delta$, much larger than $R_\star$ (i.e. when $\delta \to 0$) and start our integration of the wind from outside inward. Practically, we calculate the radiative luminosity as seen by an observer at infinity as:

$$\mathcal{L}_\infty = \dot{\mathcal{E}} - \alpha w_\infty. \tag{3.18}$$

---

[3] SUNDIALS is publicly available at `https://computation.llnl.gov/casc/sundials/main.html`.

At x_∞, the luminosity Γ_∞ comoving with the flow can be considered constant and given by:

$$\Gamma_\infty = \frac{\mathcal{L}_\infty}{1 + 2w_\infty^{1/2}/\beta} = \frac{\dot{\mathcal{E}} - \alpha w_\infty}{1 + 2w_\infty^{1/2}/\beta}. \tag{3.19}$$

We can then integrate the radiation energy density in the optically thin limit:

$$u' \simeq -2(1-x)\Gamma_\infty \quad \Rightarrow \quad u = \Gamma_\infty (1-x)^2, \tag{3.20}$$

where we use the boundary condition $u(1) = 0$, and the wind velocity:

$$w' = \Gamma_\infty - 1 \quad \Rightarrow \quad w = w_\infty - (\Gamma_\infty - 1)(1-x). \tag{3.21}$$

We can also write the explicit behaviour of $\tilde{\tau}$:

$$\tilde{\tau} = \frac{\alpha\beta(1-x)}{\sqrt{w_\infty - (\Gamma_\infty - 1)(1-x)}}. \tag{3.22}$$

Note that equation (3.19) comes out naturally by evaluating equation (3.17) at $x \to 1$, where $1+g(\tilde{\tau}) \to 2$ and u and $\tilde{\tau}$ are described by the expressions above (equations 3.20 and 3.22). Next, we choose a value for δ, typically $\delta \sim 10^{-6}$, we check that indeed $\tilde{\tau}(x_\infty) \ll 1$ and we use the formulas above to provide the initial conditions $w(x_\infty)$ and $u(x_\infty)$ for the wind equations.

We then integrate the equations inward up to the surface $x = 0$ (or up to the point where a solution exists). We define the photosphere as the place where the equality $L_{\rm phot} = 4\pi R_{\rm phot}^2 \sigma T_{\rm phot}^4$ is satisfied, where $\sigma = ca/4$ is the Stephan-Boltzmann constant, whereas the temperature T is defined from the energy density U as $T = (U/a)^{1/4}$, regardless of the local optical depth. Such a temperature is a proxy for the local temperature and it recovers its full physical meaning only when $\tilde{\tau} > 1$. The photosphere identified in this way usually lays at $\tilde{\tau} \sim 2-3$.

Every solution of the wind equations is specified by the parameters α, $\dot{\mathcal{E}}$ and w_∞, once the underlying star is set by β. Among those parameters, α is directly related to the outflow rate $\dot{M}$ and is necessary to solve the wind equations, i.e. such a model does not allow to infer theoretically the value of $\dot{M}$ a priori. However, an *acceptable* solution has to satisfy additional self-consistency requirements, which in turns impose constraints of the parameter space and ultimately on the value of $\dot{M}$. Those self-consistency requirements are imposed by the assumption that the wind originates from a star-like object. In particular: (i) the solution has to extend inward to at least $x = 0$; (ii) the wind has to be optically thick close to the surface of the star, i.e. the photosphere has to be above the base of the wind, namely $\tilde{\tau}_\star > 1$ and $R_{\rm phot} > R_\star$; and (iii) the wind has to connect to an hydrostatic solution, i.e. it should be initially subsonic (i.e. $\mathcal{M}_\star < 1$) and with a moderate velocity[4] (i.e. $w_\star \ll 1$). A wind solution is then accepted only when it satisfies all the conditions listed above, and it is discarded otherwise.

---

[4] As consequence of the assumption of steady-state, we note that we cannot allow for $w_\star = 0$ because the density would otherwise diverge, as implied by the conservation of mass in equation (3.1).

The wind velocity at infinity

To try and simplify further our procedure, we first assess the sensitivity of our solution to our choice of w_∞. The arbitrariness of w_∞ is simply a consequence of our neglecting gas pressure, in the equations describing a radiation dominated wind. When thermal gas pressure is explicitly accounted for, the sonic point of a solution (i.e. where $\mathcal{M} = 1$) is also a critical point (i.e. a divergent point for w'). The requirement on the position of the critical point to cure the local divergency translates naturally into a condition that fixes the value of w_∞. As a consequence, solutions of the wind equations with gas pressure only dependent on α and $\dot{\mathcal{E}}$ (e.g. Quinn & Paczynski, 1985). In our solutions, instead, the sonic point is *not* a critical point, and w_∞ is not univocally determined. However, the fact that with gas pressure terms there is only a single value for w_∞ and that solutions should be continuous as $p_{\rm gas}/p_{\rm rad} \to 0$ suggests that the range of possible w_∞ may be narrow. Therefore, we investigate this possibility.

We setup a grid of five representative values for $\beta \in \{10, 50, 100, 500, 1000\}$, and five representative values for $\dot{\mathcal{E}} \in \{1.5, 2.5, 5, 7.5, 10\}$. For each pair $(\beta, \dot{\mathcal{E}})$, we divide the interval $\log\alpha \in [-3, 1]$ uniformly, and for each value of α we run 10^3 integrations of the wind equations choosing a random value for w_∞ distributed uniformly in the logarithmic interval $[-3, \log(\dot{\mathcal{E}}/\alpha)]$. We keep only the acceptable solutions according to Section 3.2.2.

Our results confirm that the range of w_∞ that leads to self-consistent solutions is narrow, usually $\lesssim 0.1$ dex, and centred around $w_\infty \sim 1$. The values of w_∞ also correlate with $\mathcal{M}_\star$ in the interval $0.1 \lesssim \mathcal{M}_\star < 1$. Such a correlation is shown for 500 realisations in Figure 3.2 for an example configuration with $\beta = 100$, $\alpha = 1$ and $\dot{\mathcal{E}} = 5$ and exhibits typical features common to all the other combinations of parameters. In particular, most of the interval of allowed w_∞ corresponds to values of the Mach number larger than $\sim 0.5-0.6$, as shown by the distribution of $\mathcal{M}_\star$ represented in the inset of Figure 3.2, peaking around $\mathcal{M}_\star \sim 0.8$. Motivated by that, we can use this occurrence as an approximate additional constraint to remove the freedom of choosing w_∞ by choosing a value for $\mathcal{M}_\star$ to be matched at $R_\star$. In the following, we focus only on solutions with $\mathcal{M}_\star = 0.8 \pm 0.05$.

3.2.3 Results

We are now in the position to explore the properties of the solutions within the parameter space $(\alpha, \dot{\mathcal{E}})$ as a function of β. Figure 3.3 highlights the boundaries on the parameter space imposed by the self-consistency conditions. Solutions are limited from below by the requirement that $\dot{\mathcal{E}} \geq 1$. We assume this condition as necessary in order to launch the wind. In fact, the actual condition to have an accelerating wind is $\Gamma_\star > 1$ or in other words, that the star should shine above the Eddington limit (equation 3.12). Nonetheless, we conveniently chose the limit $\dot{\mathcal{E}} \geq 1$ because (i) it implies $\Gamma_\star > 1$ and (ii) the hydrostatic solution (the star) below the wind naturally provides $\dot{E} \approx L_\star$ as a boundary condition (we

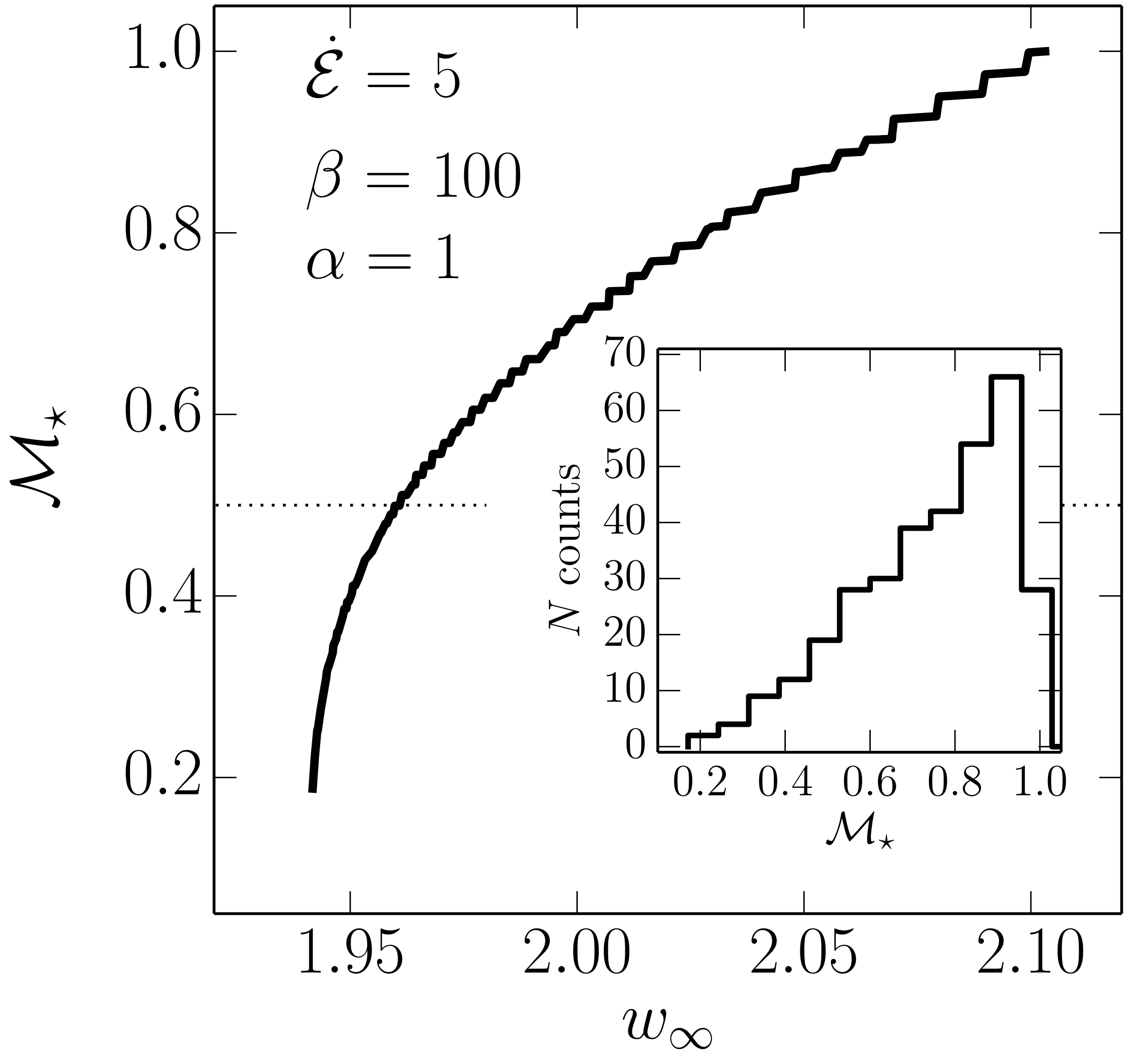

Figure 3.2 Relation between the velocity at infinity w_∞ and the Mach number at the surface $\mathcal{M}_\star$ assuming $\beta = 100$, $\alpha = 1$ and $\dot{\mathcal{E}} = 5$. The dotted thin line marks $\mathcal{M}_\star = 0.5$ for reference. The range of possible w_∞ is very narrow. The inset shows the distribution of the Mach numbers $\mathcal{M}_\star$ obtained by 500 realisations. A clear peak around $0.8 - 0.9$ is present.

discuss this point with more details in Section 3.3). The limit at small α and $\dot{\mathcal{E}} \lesssim 10$ results from imposing $\tilde{\tau}_\star > 1$ and it depends on β since the normalisation of the optical depth is $\propto \alpha\beta$, as shown by equation (3.14) and in the upper panel of Figure 3.3. Physically, this is because matter needs to be faster to escape from a more compact star, and from mass conservation ($\rho \propto \dot{M}/v$) it follows that a higher mass loss rate is required to maintain the same optical depth $\tilde{\tau}_\star > 1$. Finally, at fixed α, the upper value of $\dot{\mathcal{E}}$ is constrained by the matching with a hydrostatic solution below the wind, that requires $w_\star < 1$. Incidentally, we note here that our set of prescriptions do not set an upper limit on $\dot{E}$. This will be provided by the physical characteristics of the stellar object powering the wind, more explicitly by how much super Eddington its emission is.

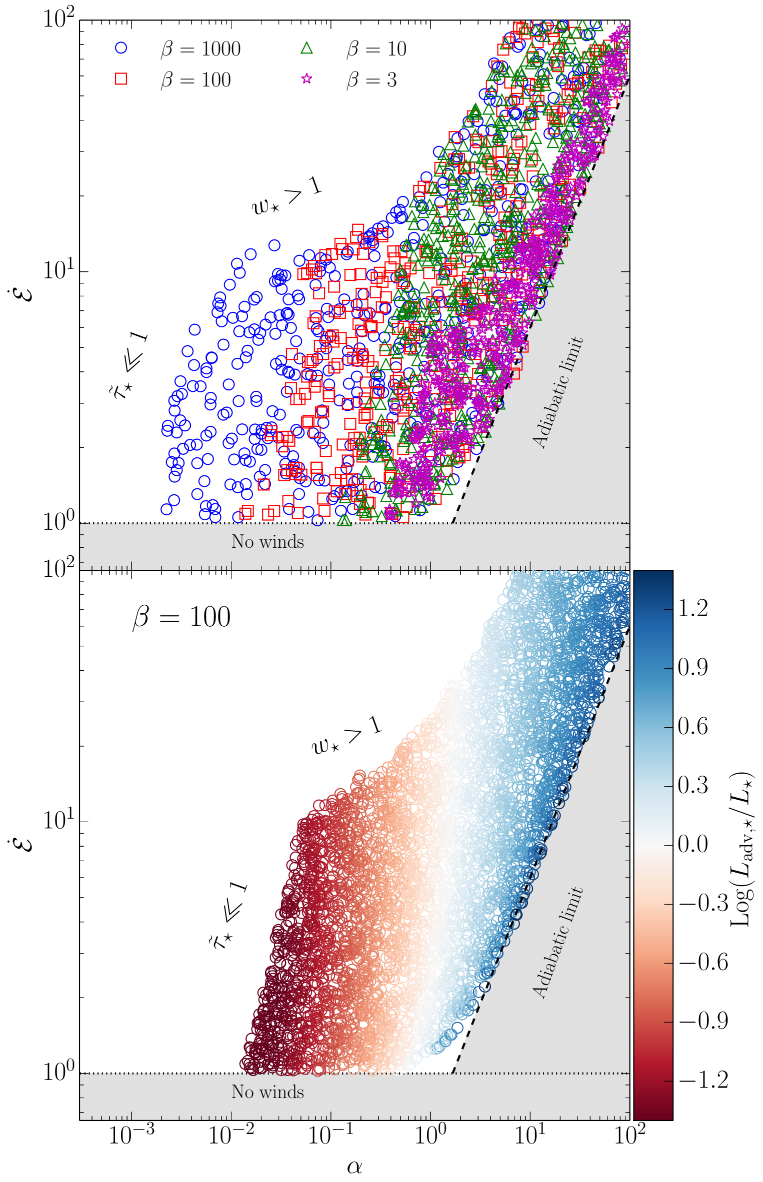

Figure 3.3 Monte Carlo sampling of the parameter space $\alpha - \dot{\mathcal{E}}$. Upper-panel: blue circles, red squares, green triangles and magenta stars show the results for $\beta = 1000$, 100, 10, and 3, respectively. For visualisation purposes, we show 10% of the 10^4 calculations performed for each value of β. Lower panel: the same as above showing all the results for the case $\beta = 100$ where we colour-coded the points according to the ratio between the advected luminosity at the surface $L_{\mathrm{adv},\star}$ and the radiative luminosity $L_\star$ at $R_\star$. Both panels show the limits to the parameter space imposed by the self-consistency conditions as described in the text.

The maximum mass loss rate for a given $\dot{\mathcal{E}}$ (grey area on the right) is instead a physical limit. This is obtained when the dominant energy source for the kinetic luminosity of the wind is the enthalpy of the gas and $\dot{\mathcal{E}} \approx w_\infty\, \alpha$. In practise, this is the behaviour of an adiabatic wind. We show this by considering a fully adiabatic solution. This latter has no radiative luminosity in its governing equations (see Appendix 3.A), and the presence of a critical point allows us to relate the velocity at infinity $w_{\infty,\mathrm{adiab}}$ to the condition at the base of the wind. We can therefore derive that an adiabatic wind with $\mathcal{M}_\star = 0.8$ will have $w_{\infty,\mathrm{adiab}} = 3w_\mathrm{c} \approx 0.6$, where w_c is the velocity at the critical point $w_\mathrm{c} = s_\mathrm{c}$, related to $\mathcal{M}_\star$ by equation (3.71) and the relation plotted in Figure 3.12. For a given $\dot{\mathcal{E}}$, we therefore have:

$$\alpha_\mathrm{max} \approx \dot{\mathcal{E}}/0.6. \tag{3.23}$$

This relation is plotted as a dashed line in Figure 3.3 and clearly marks the rightmost limit of our solutions and the beginning of the "forbidden" region dubbed "adiabatic limit".

That our solutions tend towards an adiabatic behaviour as $\dot{M}$ increases, can be better appreciated by looking at the lower panel of Figure 3.3, which shows solutions for $\beta = 100$, colour coded according to the ratio $L_{\mathrm{adv},\star}/L_\star$, where $L_\mathrm{adv} = \dot{M}(p+U)/\rho$ is the luminosity advected within the bulk outflow. Across $\alpha \sim 1$ the regime of the wind changes: for $\alpha < 1$ energy in radiation is mainly transported by diffusion while for $\alpha > 1$ advection becomes more and more dominant as the outflow rate increases towards the adiabatic limit. As mentioned before, this is exactly the physical meaning of the trapping radius (see also Figure 3.4, lower rightmost panel) and $\alpha > 1$ implies that the trapping radius occurs within the outflow, $R_\mathrm{tr} > R_\star$ (see equation 3.15). Formally, a fully adiabatic solution has $R_\mathrm{tr} = \infty$ (i.e. $\alpha = \infty$) and correspondently $L_{\mathrm{adv},\star}/L_\star = \infty$. Advection is indeed the only transport mechanism in an adiabatic wind. We can go a step further and calculate $\dot{M}$ of a solution *relative* to the maximum possible mass loss rate. The latter corresponds to that of an adiabatic wind ($\dot{M}_\mathrm{ad}$, equation 3.74) launched from the same star, with the same initial conditions (i.e. the same $c_{\mathrm{s},\star}$ and $\rho_\star$ at $R_\star$) at the same $\dot{\mathcal{E}}$ (which for an adiabatic wind reads $\dot{\mathcal{E}} = \dot{M}_\mathrm{ad}(v_\star^2/2 + 3c_{\mathrm{s},\star}^2 - GM_\star/R_\star)/L_\mathrm{Edd}$):

$$\frac{\dot{M}}{\dot{M}_\mathrm{ad}} \approx \left(1 - \frac{L_\star}{\dot{E}}\right). \tag{3.24}$$

When the contribution of the radiative luminosity $L_\star$ at $R_\star$ to the total energetic budget becomes negligible, $\dot{M}$ approaches the adiabatic value.

The upper panel of Figure 3.3 shows the comparison of 10^4 realisations for four different values of β. We find that the winds originating from more compact stars (i.e. those with smaller β) sustain higher outflow rates at the same total luminosity $\dot{\mathcal{E}}$. This is again related to the fact that gas needs to be faster to escape from a more compact star, and it compensates this increase in velocity by an increase in $\dot{M}$ to maintain optical thickness. As a consequence, an *optically-thick* wind needs to carry a larger and larger fraction of the total luminosity in kinetic and advective form. This explains why the range of possible $\dot{M}$ gets squeezed towards the adiabatic limit for[5] $\beta \to 1$.

Figure 3.4 shows two example solutions of the wind equations with $\beta = 100$. The upper row shows a solution with $\alpha = 0.4$ and $\dot{\mathcal{E}} = 5$, i.e. at the centre of the allowed region in

---

[5] Formally, such a limit cannot be reliably modelled by our wind equations, because it would imply approaching a regime where general relativistic corrections might become relevant.

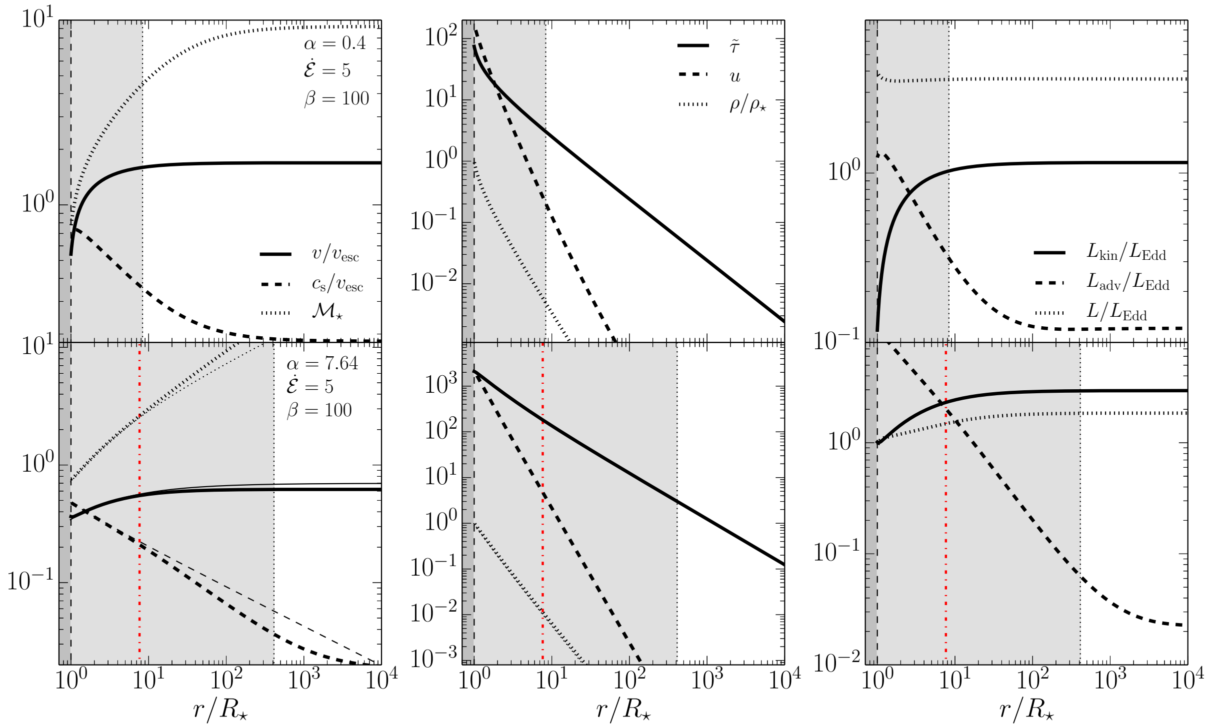

Figure 3.4 Example solutions of the wind equations for $\beta = 100$. The upper row shows the solution for $\alpha = 0.4$ and $\dot{\mathcal{E}} = 5$, while the bottom row shows the solution for $\alpha = 7.64$ and $\dot{\mathcal{E}}$, very close to the adiabatic limit of Figure 3.3. Left column: continuous, dashed and dotted lines show the profile of $v/v_{\rm esc}$, $c_{\rm s}/v_{\rm esc}$ and $\mathcal{M}$, respectively. Central column: continuous, dashed and dotted lines show the profile of $\tilde{\tau}$, u and $\rho/\rho_\star$, respectively. Right column: continuous, dashed and dotted lines show the profile of $L_{\rm kin}/L_{\rm Edd}$, $L_{\rm adv}/L_{\rm Edd}$ and $L/L_{\rm Edd}$, respectively. The dark and light grey shaded regions in all panels mark the base of the wind and the surface of the photosphere, respectively, while the red, vertical, dash-dotted lines in the bottom row indicate the position of the trapping radius. The thin lines in the bottom-left panel and the thin dotted line in the bottom-central panel show the velocity, sound speed, Mach number and density, respectively, of the adiabatic wind with the same critical point of the wind solution. The two solutions are very similar within $R_{\rm tr}$.

the lower panel of Figure 3.3, while the bottom row shows a solution with $\alpha = 7.64$ and $\dot{\mathcal{E}} = 5$, i.e. very close to the adiabatic limit. The wind velocity of the first solution grows steeply within the photosphere and then flattens to its asymptotic values; on the contrary, the sound speed $c_{\rm s}$ decreases quickly within the photosphere, matching the gas velocity at the sonic point very close to $R_\star$, beyond which the wind becomes highly supersonic. Note that, regardless of the optical depth, we always use the general definition of sound speed:

$$c_{\rm s}^2 = \left(\frac{\partial p}{\partial \rho}\right)_{\mathcal{S}} = \frac{4}{3}\frac{p}{\rho}, \tag{3.25}$$

where $\mathcal{S}$ is the specific entropy and the second equality is based on the most general equation of state for a non-isentropic, radiation-pressure dominated fluid, namely $p(\rho, \mathcal{S}) = K(\mathcal{S})\rho^{4/3}$. In fact, this relation is only valid when $p = U/3$, i.e. in the optically-thick limit, but we need to consider the behaviour of $c_{\rm s}$ only within the photosphere. The solution close to the adiabatic limit shows a similar behaviour, though the gas velocity grows less steeply than in the previous case and mostly within $R_{\rm tr}$. It also reaches an asymptotic velocity lower than the previous case, though the total energy $\dot{\mathcal{E}}$ is the same.

The profiles of the optical depth, density and radiation energy density are similar in both examples. We can fairly accurately describe them with power laws, at least close to and outside the photosphere. In particular, the optical depth decreases with radius as r^{-1}, while the density as r^{-2} once the gas velocity remains almost constant. This confirms a posteriori that $\tilde{\tau}$ is a good approximation for τ. We note also that the wind is highly optically thick close to $R_\star$, while the photosphere forms at $\tilde{\tau}_{\rm phot} \sim 2-3$. This latter confirms that our approximate treatment of the transition between the optically thin and thick regimes gives sensible results, close to the conventional $\tau_{\rm phot} \approx 1$.

The largest difference between the two solutions is in the energy budget, shown in the left column of Figure 3.4. There, we compare the profiles of the radiative luminosity L, the advected luminosity $L_{\rm adv}$, and the kinetic luminosity $L_{\rm kin} = \dot{M}v^2/2$, all normalised to $L_{\rm Edd}$ for convenience. These luminosities, summed up with $-G\dot{M}M_\star/r$, give the total, constant luminosity $\dot{E}$. The energy budget in the first solution is dominated by the radiative luminosity from the base of the wind (indeed $R_{\rm tr} < R_\star$) to infinity. After a small decrease, it remains almost constant with radius, while $L_{\rm kin}$ quickly rises within the photosphere, yet remaining subdominant. This behaviour indicates that the acceleration of the wind is not powered by L but it rather occurs at the expenses of $L_{\rm adv}$, that drops accordingly within the photosphere. In contrast, $L_{\rm adv}$ in the $\alpha = 7.64$ case (lower panel) is initially an order of magnitude higher than both L and $L_{\rm kin}$ and becomes comparable to L around $R_{\rm tr}$. Instead, L remains always subdominant compared to $L_{\rm kin}$. The radiative luminosity's mild growth within the photosphere is similar to that of an "effective" Eddington ratio that one would obtain from equation (3.12), $\Gamma_{\rm eff} = 1 + w'$, when w' from the adiabatic equations (3.68) and (3.69) is used. Notably, also in this case, $L_{\rm kin}$ is ultimately powered by $L_{\rm adv}$.

Finally, we show explicitly the similarity between the solution close to the adiabatic limit and an actual adiabatic solution in the bottom row of Figure 3.4. Specifically, after calculating the solution of our wind equations, we calculate also the adiabatic solution (see Appendix 3.A) which has the same critical radius. Such an adiabatic solution crosses naturally the critical point with the same value of critical velocity $w_{\rm c}$ of the full solution, and closely resembles it (comparing v, $c_{\rm s}$, $\mathcal{M}$, and ρ) within the trapping radius, outside of which radiative diffusion becomes relevant. From our results, it is clear that the ultimate source of kinetic energy for the gas is advection energy and *not* the diffusive radiation luminosity. This is not consumed as the gas propagates outward. To increase the mass loss, it is therefore necessary to go towards an adiabatic solution, where initially the advection energy dominates the energy budget. This is different from the results by Shaviv (2001b), Owocki, Gayley & Shaviv (2004), i.e. a "photon-tired" wind.

3.3 Winds from quasi-stars

Directly forming massive black hole seeds is possible in principle when a mass of $\sim 10^8$ $\mathrm{M_\odot}$ can be collected in $\sim 10^6$ yr, requiring inflow rate > 100 $\mathrm{M_\odot}$ $\mathrm{yr^{-1}}$. The reason is that such a rapid accumulation of mass has to occur before nuclear reactions dominate the evolution, setting a lifetime of $\sim 10^6$ yr (Begelman, 2010). In these extreme conditions, a dynamical instability due to relativistic effects may develop and even rotation cannot prevent the cloud from collapsing directly into a black hole (Fowler 1966; Baumgarte & Shapiro 1999; Shibata & Shapiro 2002; see however Ferrara et al. 2014). On the other hand, milder conditions (e.g. inflow rate $\sim 0.1 - 1$ $\mathrm{M_\odot}$ $\mathrm{yr^{-1}}$) would lead to the formation of a supermassive star, possibly $\sim 10^5 - 10^7$ $\mathrm{M_\odot}$, stabilised by some rotation (Begelman, 2010; Hosokawa, Omukai & Yorke, 2012; Hosokawa et al., 2013). After $\sim 10^6$ yr, a small embryo seed ($\lesssim 100$ $\mathrm{M_\odot}$) can form at the centre of a supermassive star at the end of the hydrogen-burning phase (Begelman, 2010). Such a seed then needs to go through a phase of vigorous super-Eddington accretion to reach $10^4 - 10^5$ $\mathrm{M_\odot}$ within a few million years from its birth. This super-Eddington accretion can occur within a quasi-star: a very massive ($> 10^5 - 10^6$ $\mathrm{M_\odot}$) quasi hydrostatic envelope that surrounds the black hole and feeds it at a rate equal to roughly its own (i.e. quasi-star's) Eddington limit (Begelman, Rossi & Armitage, 2008). Accretion around the Eddington limit involves radiation dominated gas, which is loosely bound with a total energy close to zero. For this reason outflows can easily form. Here we investigate whether this vigorous accretion in quasi-stars is also accompanied by mass loss, as expected in other super-Eddington systems such as discs (Blandford & Begelman, 2004).

3.3.1 Equations

We follow Begelman, Rossi & Armitage (2008) and Dotan, Rossi & Shaviv (2011) to describe the hydrostatic envelope of a quasi-star and then we match it with our wind model, looking for equilibrium solutions. A quasi-star is made of four components: (i) the central, accreting black hole, (ii) a convective, radiation-pressure dominated envelope, (iii) a porous radiative layer, and (iv) a wind. In the following, we briefly describe components i-iii, remanding to Dotan, Rossi & Shaviv (2011) for additional details. We do not model explicitly the inflow on to the central black hole; instead, we treat it as a boundary condition, assuming that a black hole of mass $M_\bullet$ is accreting through a convection-dominated disc (Stone, Pringle & Begelman, 1999; Igumenshchev & Abramowicz, 1999; Quataert & Gruzinov, 2000; Agol et al., 2001) within a few Bondi radii $r_\mathrm{B} = GM_\bullet/(2c_\mathrm{s,c}^2)$, where $c_\mathrm{s,c}$ is the central sound speed $\lesssim r_\mathrm{B}$. We assume that the black hole is radiating at a luminosity $L_\bullet$ close to the maximum that convection-dominated accretion flows can sustain evaluated at $5r_{\rm B}$, namely:

$$L_\bullet = L_{\rm conv}(5r_{\rm B}) = 4\ \pi\ (5r_{\rm B})^2\ \rho_{\rm c}\ c_{\rm s,c}^3, \tag{3.26}$$

where $\rho_{\rm c}$ is the central density outside a few $r_{\rm B}$. $L_\bullet$ it is injected at the centre of the envelope and transported till the base of the wind, first by convection and then by diffusion. We used it as an inner boundary condition for the integration of the envelope and we also neglect the gas mass within $5r_{\rm B}$, assuming that only $M_\bullet$ contributes at smaller radii. We checked the effect of varying the position of the inner boundary. We find that most of the properties in the $M_\bullet - M_\star$ plane (see Section 3.3.3) remain unchanged. However, as discussed by Ball et al. (2011) and Ball, Tout & Żytkow (2012), changing the inner radius modifies (at the same central pressure) the ratio $M_\bullet/M_\star$ of two consistent solutions. This slightly displaces the no-hydrostatic-solution region. Although a few Bondi radii are a reasonable estimate for the inner accretion region where the black hole gravity is expected to dominate, we caution that such a choice remain somewhat arbitrary.

Outside $5r_{\rm b}$, the radiation-pressure dominated, convective envelope extends; we assume that it satisfies the hydrostatic equilibrium:

$$\frac{{\rm d}P}{{\rm d}r} = -\frac{GM(r)\rho}{r^2}, \tag{3.27}$$

where the total pressure P is the sum of the gas pressure $P_{\rm g} = \rho k_{\rm B}T/(\mu m_{\rm p})$ and the radiation pressure $P_{\rm r} = aT^4/3$, specified by the gas density ρ and temperature T. $k_{\rm B}$, a and $m_{\rm p}$ are the Boltzmann constant, the radiation constant and the mass of the proton, respectively. We assume the mean molecular weight $\mu = 0.59$, appropriate for gas with primordial composition at $T > 10^4$ K, as usually true everywhere in the interior of quasi-stars. The enclosed mass $M(r)$ is given by:

$$M(r) = M_\bullet + 4\pi \int_{5r_{\rm B}}^{r} \rho(r')r'^2{\rm d}r'. \tag{3.28}$$

The dominant energy transport mechanism within the envelope is convection, which induces a temperature gradient very close to adiabatic; therefore, we evolve the temperature gradient assuming that it is equal to the adiabatic one:

$$\frac{{\rm d}\log T}{{\rm d}\log P} = \frac{\gamma_{\rm ad} - 1}{\gamma_{\rm ad}}, \tag{3.29}$$

where the adiabatic index $\gamma_{\rm ad}$ depends on the ratio $\zeta \equiv P_{\rm g}/P$ according to:

$$\gamma_{\rm ad} = \frac{32 - 24\zeta - 3\zeta^2}{24 - 18\zeta - 3\zeta^2}. \tag{3.30}$$

The convective envelope extends till the radius $r_{\rm conv}$ where $L_{\rm conv}(r_{\rm con}) = L_\bullet$. Outside $r_{\rm conv}$, $L_{\rm conv} < L_\bullet$ and diffusion becomes more efficient in transporting $L_\bullet$ than convection. However, $L_\bullet$ may be larger than the local Eddington limit associated to the local enclosed mass $M(r)$ and to electron scattering opacity. In such a condition, the gas becomes locally unstable and develops inhomogeneities (Shaviv, 2001a) that have the effect of reducing the effective opacity with respect to its microscopic value even before than the Eddington limit is reached. Following Dotan, Rossi & Shaviv (2011), we model the effective opacity κ_{eff} as:

$$\kappa_{\text{eff}} = \begin{cases} \dfrac{\kappa}{\Gamma}\left(1 - \dfrac{0.16}{\Gamma}\right) & \Gamma > 0.8, \\ \kappa & \Gamma \leq 0.8, \end{cases} \tag{3.31}$$

where $\Gamma = L_\bullet / L_{\text{Edd}}$ is the local Eddington ratio calculated using equation (3.11) with $M(r)$, while κ is the microscopic opacity:

$$\kappa(T) = \frac{\kappa_{\text{es}}}{1 + (T/T_0)^{-13}}, \tag{3.32}$$

where $T_0 = 8000$ K. This opacity models the results for pristine gas by Mayer & Duschl (2005). The effective opacity corresponds to an effective Eddington ratio $\Gamma_{\text{eff}} = 1 - 0.16/\Gamma$ when $\Gamma > 0.8$, i.e. the gas is effectively sub-Eddington, though it would be super-Eddington with the microscopic opacity. Throughout this radiative layer (that usually encompasses a tiny fraction of the total mass), we assume once again hydrostatic equilibrium (being effectively sub-Eddington) and we solve equations (3.27) and (3.28), but we evolve the temperature by mean of the radiative gradient with κ_{eff}:

$$\frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{3\kappa_{\text{eff}}\rho L_\bullet}{16\pi acr^2T^3}. \tag{3.33}$$

The luminosity $L_\bullet$ remains constant since no energy sources/sinks are present within the convective envelope or the radiative layer. The inhomogeneities in the radiative layer can maintain the luminosity sub-Eddington as long as they remain optically thick. Since those inhomogeneity have a size of order of the local density scale-height, we can estimate their optical depth as $\tau_{\text{eff}} \approx \chi\rho\kappa_{\text{eff}}h$, where χ is the ionisation fraction calculated from the Saha equation assuming equilibrium, and $h = |\rho/(\mathrm{d}\rho/\mathrm{d}r)|$ is the density scale-height. Then, the radiative layer extends up to r_{rad} such that $\tau_{\text{eff}}(r_{\text{rad}}) = 1$. We note that ζ typically decreases quickly throughout the radiative layer, reaching values $\zeta \ll 0.01$ at r_{rad}.

We can finally solve and connect the wind model described in Section 3.2. In particular, we use $M_\star = M(r_{\text{rad}})$ and $R_\star = r_{\text{rad}}$. From these quantities we can evaluate β associated to the star. As in Section 3.2, we do not model the initial acceleration of mass explicitly. Instead, we assume that this occurs very quickly around r_{rad}, which represent the interface between the hydrostatic part and the wind. Then, we assign $\dot{E} = L_\bullet$ to guarantee the conservation of energy at the interface, because below r_{rad} there is no net displacement of mass and the total luminosity transported is just $L_\bullet$. Finally, we need to specify $\dot{M}_{\text{wind}}$. As described in Section 3.2.3, we assume that the wind connect with the hydrostatic part with a fixed $\mathcal{M}_\star < 1$. At the same time, we assume continuity for the density and pressure at r_{rad}, which implies:

$$\dot{M}_{\text{wind}} = 4\ \pi\ r_{\text{rad}}^2 \rho_\star \mathcal{M}_\star c_\star, \tag{3.34}$$

where $\rho_\star = \rho(r_{\rm rad})$ and $c_\star = c_{\rm s}(r_{\rm rad})$ are the density and the sound speed evaluated at $r_{\rm rad}$ as given by the integration of the radiative layer, respectively. Once we have β, $\dot{E}$ and $\dot{M}$, we can integrate the wind equations as described in Section 3.2.2 and 3.2.3, with the only difference that we use the temperature-dependent opacity of equation (3.32). We evaluate it using as a proxy for the local temperature $T = (U/a)^{1/4}$, which is correct only in the optically-thick part of the atmosphere.

3.3.2 Numerical integration

We proceed to describe the numerical strategy to solve the equations in Section 3.3.1, similar to what has been done by Dotan, Rossi & Shaviv (2011).

1. We choose the black hole mass $M_\bullet$.
2. We choose one of the central quantity, in particular the total central pressure $P_{\rm c}$.
3. We need a second quantity to specify all the boundary conditions at the centre. Therefore, we guess the value of $\zeta_{\rm c} = P_{\rm gas,c}/P_{\rm c}$.
4. We calculate the central quantities: $T_{\rm c} = 3(1-\zeta_{\rm c})P_{\rm c}/a$, $\rho_{\rm c} = P_{\rm c}\zeta_{\rm c}\mu m_{\rm p}/(k_{\rm B}T_{\rm c})$, $c^2_{\rm s,c} = ((4/3)(1-\zeta_{\rm c})+(5/3)\zeta_{\rm c})P_{\rm c}/\rho_{\rm c}$ and $r_{\rm B} = GM_\bullet/(2c^2_{\rm s,c})$; we evaluate $L_\bullet$ according to equation (3.26).
5. We integrate the convective envelope, namely equations (3.27), (3.28) and (3.29), from the centre ($5r_{\rm B}$) outward, until we reach $r_{\rm conv}$. We actually integrate the equation in their Lagrangian form, using the enclosed mass $M(r)$ as the independent variable.
6. We integrate the radiative layer equations (3.27), (3.28) and (3.33), assuming continuity with the convective envelope from $r_{\rm conv}$ till $r_{\rm rad}$. We integrate the equations using P as the independent variable since it varies more than the other quantities throughout the radiative layer.
7. We calculate the necessary quantities to specify the properties of the wind using the values at $r_{\rm rad}$ as discussed in Section 3.3.1, namely β, $\dot{E}$ and $\dot{M}_{\rm wind}$; specifically, we assume $\mathcal{M}_\star = 0.8$.
8. We calculate the wind solution and we check that/whether it reaches $r_{\rm rad}$ self-consistently as described in Section 3.2.2.
9. We compare the density $\rho^{\rm (rad)}(r_{\rm rad})$ obtained at the end of the integration of the radiative layer with the density $\rho^{\rm (wind)}(r_{\rm rad})$ obtained by the wind integration and we modify $\zeta_{\rm c}$ in order to match the two values.

We find empirically that the ratio $\rho^{(\mathrm{wind})}(r_{\mathrm{rad}})/\rho^{(\mathrm{rad})}(r_{\mathrm{rad}})$ crosses the value 1 extremely steeply while varying ζ_{c} and is not monotone far from the solution. This occurrence makes difficult to use classic methods such as bisection unless the initial guesses for the values of ζ_{c} that bracket the final solution are very close to the latter. To overcome this problem, we proceed as follow: we choose an initial guess for ζ_{c} by solving the following equation that comes from the scaling relations of the envelope provided by Begelman, Rossi & Armitage (2008) and Dotan, Rossi & Shaviv (2011):

$$p_{\mathrm{c},7} = \frac{1.134}{m_{\bullet}^{16/5}} \frac{(1-\zeta_{\mathrm{c}})^{2/5}}{\zeta_{\mathrm{c}}^{28/5}}, \tag{3.35}$$

where $P_{\mathrm{c}} = p_{\mathrm{c},7} \times 10^{7}$ erg cm^{-3} and $M_{\bullet} = m_{\bullet}$ M$_{\odot}$. Then, we build a grid of models for several values of ζ_{c} around the initial guess, and we progressively refine this grid around the solution. When we bracket the true solution with a relative precision $\sim 10^{-3}$, we use this bracketing as the starting points for a Brent root-finder (Brent, 1973; Press et al., 2002). The typical solutions of ζ_{c} are $\lesssim 0.01$; indeed, quasi-stars are radiation-pressure dominated in their interiors. Moreover, we note that ζ typically decreases throughout the radiative layer, reaching values $\zeta \ll 0.01$ at r_{rad}. This behaviour justifies our simplifying assumption of neglecting the gas pressure in the wind, since the ratio $3k_{\mathrm{B}}\rho/(\mu m_{\mathrm{p}} a T^{3})$ remains effectively $\ll 1$ through the wind and within the photosphere. Calculations a posteriori of ζ in the wind show a decreasing behaviour. This suggests that our treatment is at least consistent. Of course, the radiation dominated assumption limits our results to very massive stars and cannot be extended to e.g. Population III stars.

3.3.3 Results for quasi-stars

We run a grid of models exploring a wide range of $M_{\bullet}$ and P_{c}, which maps into $M_{\star}$. Figure 3.5 summarises our findings in the $M_{\bullet} - M_{\star}$ plane. Specifically, the various panel shows iso-contours of the outflow rate $\dot{M}_{\mathrm{wind}}$, the accretion onto the black hole $\dot{M}_{\mathrm{BH}} = L_{\bullet}/(\eta c^{2})$, where the radiative efficiency $\eta \simeq 0.1$, the photospheric luminosity L_{phot}, and the effective, photospheric temperature T_{phot}.

Such a plane is characterised by three regions. For high black hole masses and relatively low envelope masses, no hydrostatic solution can be found (see also Begelman, Rossi & Armitage, 2008). That is because the quasi-stars would stay beyond the Hayashi (1961) track, which represent a lower limit to the effective temperature (around 4000 K) of a convective envelope in hydrostatic equilibrium. Beyond such a limit, no solutions for the hydrostatic envelope exist. The second region is the evaporation strip identified by Dotan, Rossi & Shaviv (2011). This region lays where the evaporation timescale $t_{\mathrm{wind}} = M_{\star}/\dot{M}_{\mathrm{wind}}$, i.e. the typical timescale over which the stellar envelope would be blown away by the winds, is shorter than the accretion timescale $t_{\bullet} = M_{\bullet}/\dot{M}_{\mathrm{BH}}$. Within this region, the envelope mass is dispersed by the wind before than the black holes can

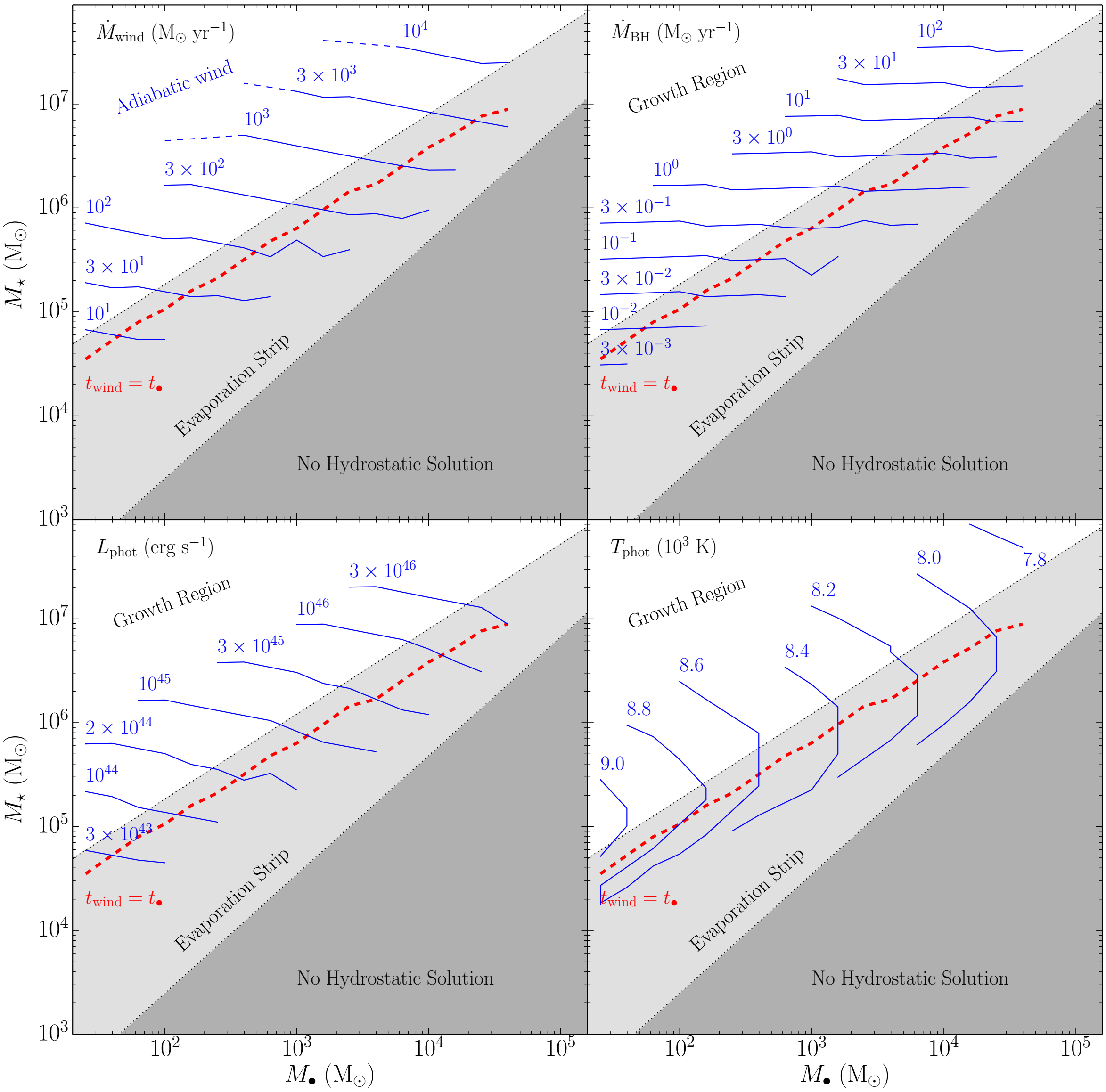

Figure 3.5 Properties of our quasi-star models in the $M_\bullet - M_\star$ plane. From upper-left panel, clockwise: thin, blue continuous lines show the isocontours of the outflow rate $\dot{M}_{\rm wind}$, of the accretion rate on to the black hole $\dot{M}_\bullet$, of the photospheric bolometric luminosity $L_{\rm phot}$, and of the photospheric effective temperature $T_{\rm phot}$, respectively. The blue, dashed lines in the upper-left panel show the extension in the adiabatic wind regime of some isocontours as discuss in the text. The adiabatic wind region is labelled explicitly in this panel, while the label “Growth Region” is omitted for clarity. In each panel, the white, light grey and dark grey shaded regions represent the growth region, the evaporation strip (i.e. where $t_{\rm wind} < t_\bullet$), and the region where no hydrostatic solutions exist as reported by Dotan, Rossi & Shaviv (2011), respectively; on the other hand, the thick, red, dashed line shows the limit of the evaporation strip (i.e. where $t_{\rm wind} = t_\bullet$) for our models.

accrete further. The third region is the growth region, where the black hole can accrete substantial mass from the envelope ($t_\bullet < t_{\rm wind}$).

We find that the iso-contours of $\dot{M}_{\rm wind}$ due to our wind model of Section 3.2 are different from those in Dotan, Rossi & Shaviv (2011). Most notably, we do not see any effect of photon-tiring, consistently with our result of Section 3.2.3 that the ultimate source of the wind kinetic luminosity is the advection luminosity due to internal energy. Specifically, our contours increase mildly when $M_\bullet$ decreases, while their would become suddenly much more steep at the onset of photon-tiring, with $\dot{M}_{\rm wind}$ almost independent of $M_\star$. At constant $M_\bullet$, the different shapes of the iso-contours are such that $\dot{M}_{\rm wind}$ grows faster at high values of $M_\star$ until it reaches the adiabatic limit discussed in Section 3.2.3.

When our models hit the adiabatic limit, our quasi-star structures are all characterised by $s_\star = c^2_{{\rm s},\star}/v^2_{\rm esc}$ very close to the value associated to $\mathcal{M}_\star \simeq 0.8$ for an adiabatic wind, as shown by Figure 3.12 in the Appendix 3.A. Above this limit, our choice of a specific surface Mach number (≈ 0.8) over-constrains our mathematical system and solutions cannot be found. Technically, we would need to leave $\mathcal{M}_\star$ free to vary and the likely result would be solutions with nearly adiabatic winds.

To test this expectation, we then change method and consider the idealised case of a purely adiabatic wind solution (see Appendix 3.A) and match it to the hydrostatic envelope. These winds are completely specified by the values of the density $\rho(r_{\rm rad})$ and of the sound speed $c_{\rm s}(r_{\rm rad})$ at the end of the envelope. We then check that s associated to $c_{\rm s}(r_{\rm rad})$ is in the interval $1/6 < s < 1/4$ in order to have a solution with $\mathcal{M}_\star < 1$ (Figure 3.12). Finally, we calculate the adiabatic mass loss $\dot{M}_{\rm ad}$ according to equation (3.74) and we choose the equilibrium solution (i.e. the value of $\zeta_{\rm c}$) that allows us to match the luminosity $L_\bullet$ to the luminosity carried by the adiabatic wind evaluated at $R_\star = r_{\rm rad}$, namely $3\dot{M}_{\rm ad}c^2_{{\rm s},\star}(1 + \mathcal{M}^2_{\rm ad}(c_{{\rm s},\star}) - v^2_{\rm esc}/(6c^2_{{\rm s},\star}))$. We find a fairly smooth transition between the two kinds of models, confirming the nearly adiabatic nature of the expected wind when this limit is exceeded (see dashed lines in Figure 3.5, top left panel).

This difference has also the effect of displacing slightly the interface between the evaporation strip and the growth region, i.e. the "threshold-growth line" (Dotan, Rossi & Shaviv, 2011). Our models predict that such a line moves by a factor of $\sim 2-3$ toward higher $M_\bullet$, decreasing the thickness of the evaporation strip. Moreover, for every point $(M_\bullet, M_\star)$ around and within the evaporation strip, we find an $\dot{M}_{\rm wind}$ smaller by a factor $\lesssim 10$ compared to Dotan, Rossi & Shaviv (2011), which implies a less sudden removal of the envelope when the quasi-stars enters the evaporation strip. On the other hand, the iso-contours of $\dot{M}_{\rm BH}$ are almost independent of $M_\bullet$ and they are similar to previous findings.

The absence of a photon-tired wind has a strong impact on the photospheric luminosity of the quasi-stars. Since the wind is mostly accelerated at the expense of the internal energy, the diffusive luminosity coming out at the photosphere $L_{\rm phot}$ is a large fraction of the luminosity $L_\bullet$ originally produced by central accretion and transported through the hydrostatic envelope. Such a luminosity corresponds to Eddington ratios calculated with respect of the whole envelope mass that range from $\Gamma_{\rm phot} \sim 1$ to $\Gamma_{\rm phot} \lesssim 10$, even in the growth region, where Dotan, Rossi & Shaviv (2011) find a decrease of $\Gamma_{\rm phot}$ due to photon-tiring. Our models predict photospheric luminosities in the interval $10^{43} \lesssim L_{\rm phot}/({\rm erg~s^{-1}}) \lesssim 10^{47}$, with iso-contours similar in shape to those of $\dot{M}_{\rm wind}$ in the $M_\bullet - M_\star$ plane. Such luminosities are comparable to moderate bolometric luminosities of quasars (e.g. Hopkins, Richards & Hernquist 2007; Mortlock et al. 2011) and might be observable at high redshift as discussed in Section 3.3.4 below. At the same time, all our models fall in a narrow range of photospheric temperature between $\sim$ 7500 K and $\sim$ 9000 K, with temperatures that decrease approaching the no-hydrostatic-solution region. We recall from Section 3.2.2 that we define the photosphere as the place where the effective temperature $T_{\rm eff} \equiv (L/(4\pi r^2\sigma))^{1/4}$ equals the proxy for the temperature $T \equiv (U/a)^{1/4}$. This happens self-consistently at relatively large optical depth $\tilde{\tau} \gtrsim 10$, where T is physically motivated and correctly influences the optical depth through the opacity law. Moreover, we explicitly check that $T_{\rm eff}$ computed where $\tilde{\tau} = 1$ changes by at most ≈ 300 K for all our models, suggesting that our determination of the effective, photospheric temperature $T_{\rm phot}$ is anyway robust. The narrow range of effective temperature is mostly set by the microphysical properties of the gas, specifically by the adopted opacity law. Indeed, the steep temperature dependence of equation (3.32) is such that the the wind becomes optically thin near the opacity drop around $T_0 = 8000$ K. However, the use of the temperature-dependent opacity law (equation 3.32) in the wind of the quasi-star models does not change significantly the main physical properties of the wind described in Section 3.2 (i.e. the behaviour of the wind when approaching the adiabatic limit), especially within the photosphere where $\kappa_{\rm eff}(T) \sim \kappa_{\rm es}$.

The iso-contours of $\dot{M}_{\rm wind}$, $\dot{M}_{\rm BH}$ (or $L_\bullet$), and $L_{\rm phot}$ behave smoothly enough in the $M_\bullet$-$M_\star$ that they can be reasonably well fitted with power-laws. By means of a least-square fitting procedure in log-space, we find the following fitting formulas:

$$\dot{M}_{\rm wind} = (1.4 \pm 0.1) \times 10^{-4}\ m_\star^{0.96}\ m_\bullet^{0.17}\ {\rm M_\odot~yr^{-1}}, \tag{3.36}$$

$$\dot{M}_{\rm BH} = (8.3 \pm 0.1) \times 10^{-10}\ m_\star^{1.45}\ m_\bullet^{0.03}\ {\rm M_\odot~yr^{-1}}, \tag{3.37}$$

$$L_{\rm phot} = (3.7 \pm 0.1) \times 10^{38}\ m_\star^{0.94}\ m_\bullet^{0.29}\ {\rm erg~s^{-1}}. \tag{3.38}$$

These formulas represent the interpolation between the models of our grid. Assuming that quasi-stars evolve through a sequence of equilibrium states (Begelman, Rossi & Armitage, 2008; Ball et al., 2011; Dotan, Rossi & Shaviv, 2011), we can use them to calculate evolutionary tracks in the $M_\bullet - M_\star$ plane. We show a few example solutions in Figure 3.6, where we assume equations (3.36) and (3.37) to be valid also in the adiabatic regime (only relevant for the most massive quasi-star in these examples). Specifically, we solve the equations $\dot{M}_\star = \dot{M}_{\rm in} - \dot{M}_{\rm wind} - \dot{M}_{\rm BH}$, and $\dot{M}_\bullet = \dot{M}_{\rm BH}$, where we allow for smooth constant accretion on to the quasi-star envelope through the constant $\dot{M}_{\rm in}$. When the quasi-star enters the evaporation strip, winds start to become dominant and the envelope mass drops while the black hole cannot grow very efficiently. However, if accretion is

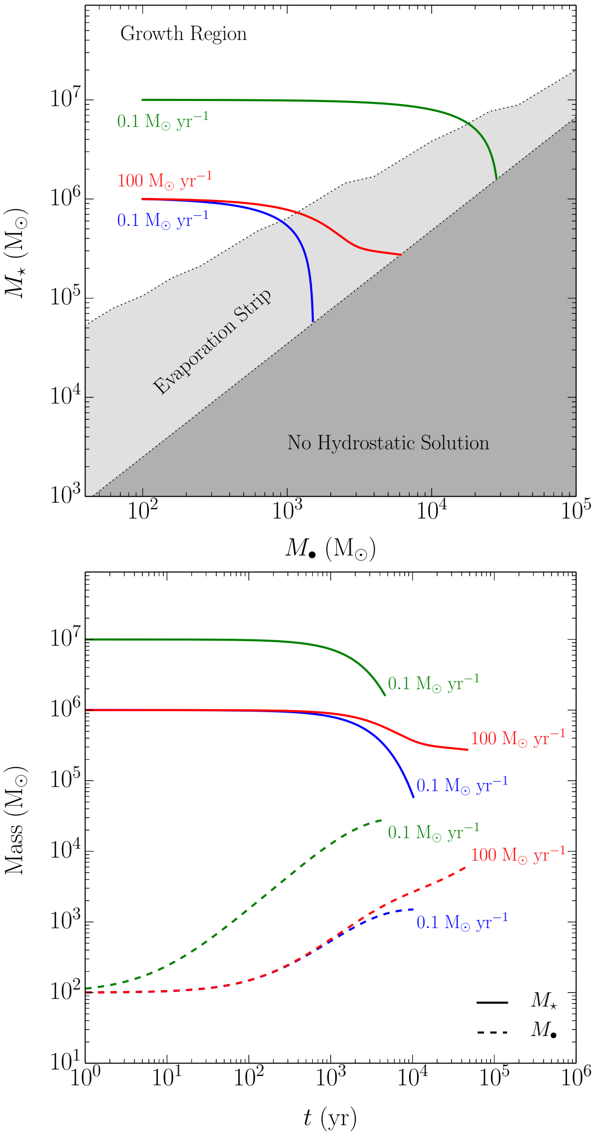

Figure 3.6 Evolutionary tracks for quasi-stars. Upper panel: evolutionary tracks in the $M_\bullet - M_\star$ plane (see Figure 3.5 for a description, but note that here we plot the evaporation strip determined from our models). The green line corresponds to a quasi-stars with initial $M_\bullet = 100\ \mathrm{M_\odot}$, $M_\star = 10^7\ \mathrm{M_\odot}$, and $\dot{M}_{\rm in} = 0.1\ \mathrm{M_\odot\ yr^{-1}}$. The blue and red lines both correspond to a quasi-star with initial $M_\bullet = 100\ \mathrm{M_\odot}$, $M_\star = 10^6\ \mathrm{M_\odot}$, but with $\dot{M}_{\rm in} = 0.1\ \mathrm{M_\odot\ yr^{-1}}$ and $\dot{M}_{\rm in} = 100\ \mathrm{M_\odot\ yr^{-1}}$, respectively. Lower panel: evolution of $M_\bullet$ and $M_\star$ as a function of time. Continuous and dashed lines (coupled with colours) show the evolution of $M_\star$ and $M_\bullet$, respectively.

intense enough, an almost steady state can be established within the evaporation strip, where the accretion from outside balances the mass loss due to winds and the black hole can grow until the quasi-stars enters the no-hydrostatic-solution region, which corresponds to a ratio $M_\star/M_\bullet \lesssim 20$.

3.3.4 Detectability

We have shown that our models of quasi-stars shine with a bolometric photospheric luminosity that is bracketed between $\sim 10^{44}$ and $\sim 10^{46}$ erg s^{-1} for $M_\star$ between $\sim 10^5$ and $\sim 10^7$ M$_\odot$. At the same time, we have found that the interval of photospheric temperature is quite narrow around 8500 K. This information allows us to put simple constraints on the detectability of those sources by current and future space-based telescopes such as Hubble Space Telescope (HST) and James Webb Space Telescope (JWST), respectively.

Indeed, we can estimate the flux in the filter band X as the magnitude:

$$m_X = M_X + d(z_{\rm source}) + K_X(z_{\rm source}), \tag{3.39}$$

where M_X is the absolute magnitude in band X, $d(z)$ is the distance modulus at the redshift $z_{\rm source}$ of the source, and $K_X(z_{\rm source})$ is the K-correction (e.g. Hogg et al. 2002). The distance modulus is defined as:

$$d(z) = 25 + 5\log_{10}\left(\frac{D_{\rm L}(z)}{\rm Mpc}\right), \tag{3.40}$$

and it encapsulates all the dependences on the cosmology through the luminosity distance $D_{\rm L}(z)$:

$$D_L(z) = (1+z)\ c\ H_0^{-1}\int_0^z \frac{{\rm d}z'}{\sqrt{\Omega_{\rm m}(1+z')^3+\Omega_\Lambda}}. \tag{3.41}$$

Here and in the following we assume the present day values $H_0 = 67.7$ km s^{-1} Mpc^{-1}, $\Omega_{\rm m} = 0.309$, and $\Omega_\Lambda = 0.691$ for the Hubble parameter, the matter density, and the density of the cosmological constant, respectively. These values are consistent with the latest Plank cosmology (Planck Collaboration et al., 2015). The absolute magnitude is the flux as if the source were 10 pc away from the observer:

$$M_X = -2.5\log_{10}\left[\frac{\displaystyle\int_0^{+\infty}\frac{L_\nu\ T_X(\nu)}{4\pi(10\ {\rm pc})^2}\ \frac{{\rm d}\nu}{\nu}}{\displaystyle\int_0^{+\infty} g_\nu\ T_X(\nu)\ \frac{{\rm d}\nu}{\nu}}\right], \tag{3.42}$$

where L_ν is the intrinsic spectral luminosity density (i.e. $L_\nu = {\rm d}L/{\rm d}\nu$) of the source, $T_X(\nu)$ is the probability of a photon to get counted at frequency ν with the filter X, and $g_\nu = 3631$ Jy is the constant spectral flux density of a hypothetical reference source used to express magnitudes in the AB system (Oke & Gunn, 1983). Following Hogg et al. (2002), we write the K-correction as:

$$\begin{aligned} K_X(z) &= -2.5\log_{10}(1+z) \\ &\quad -2.5\log_{10}\left[\frac{\displaystyle\int_0^{+\infty} L_\nu\ T_X\left(\frac{\nu}{1+z}\right)\frac{{\rm d}\nu}{\nu}}{\displaystyle\int_0^{+\infty} L_\nu\ T_X(\nu)\frac{{\rm d}\nu}{\nu}}\right]. \end{aligned} \tag{3.43}$$

The crucial ingredient is the spectra luminosity density L_ν of the source. We assume that L_ν can be modelled as a black body at the temperature $T_{\rm phot}$, emitting the total luminosity $L_{\rm phot}$. Explicitly, we have:

$$L_\nu = \frac{15 L_{\rm phot}}{\pi^4 \nu_{\rm th}} \frac{(\nu/\nu_{\rm th})^3}{\exp(\nu/\nu_{\rm th}) - 1}, \tag{3.44}$$

where $\nu_{\rm th} = k_{\rm B} T_{\rm phot}/h$ and h is the Plank constant.

Having an effective temperature ~ 8000 K, our quasi-star models are expected to be fairly blue; on the other hand, cosmologically-motivated calculations predict that quasi-stars populate mostly massive haloes at $z \gtrsim 10$ (Volonteri & Begelman, 2010), with the consequence of displacing the bulk of quasi-star emission in the near-infrared wavelengths $\gtrsim 2$ μm. Therefore, we focus our analysis on the wide filters in the near-infrared at the longest wavelength and contemporary with the highest (effective or predicted) sensibility available for HST and JWST. Specifically, we consider the filters F160W of the WFC3 camera[6] mounted on HST and the filter F444W of the NIRCam[7] designed for JWST.

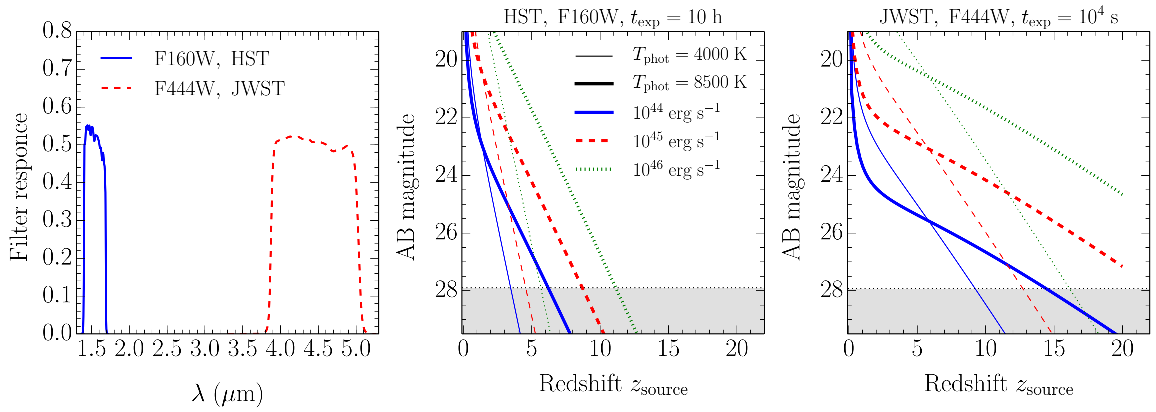

Figure 3.7 Predictions of the observability of quasi-stars. Left panel: blue, continuous line and red, dashed line show the probability that a photon with wavelength λ is captured by the WFC3 camera+F160W filter mounted on HST and by the NIRCam camera+F444W filter planned for JWST, respectively. Central panel: predicted AB magnitude in the band of the HST filter F160W as function of the source redshift $z_{\rm source}$. Blue-continuous, red-dashed and green-dotted lines refer to the photospheric luminosity $L_{\rm phot} = 10^{44}$, 10^{45}, 10^{46} erg s^{-1}, respectively; thin and thick lines refer to the effective temperature $T_{\rm phot} = 4000$, 8000 K, respectively. Right panel: the same as the central panel, but for the NIRCam camera+F444W filter planned for JWST.

The left panel of Figure 3.7 shows the throughput of the considered filters. The HST data consider also the coupling between the camera and the filter, while the JWST data are the predicted transmittance of the filter only. In order to mimic the effect of the coupling with the camera for JWST as well, we conservatively multiply the filter transmission by the fudge factor 0.6, obtaining a maximum response similar to the HST values ~ 0.5.

---

[6] `http://svo2.cab.inta-csic.es/svo/theory/fps3/`

[7] `http://www.stsci.edu/jwst/instruments/nircam/instrumentdesign/filters/`

The central and the left panel of Figure 3.7 show the predicted flux observed by HST and JWST, respectively. We explore the effect of changing the total luminosity of the quasi-star and its effective temperature $T_{\rm phot}$ to compare with Volonteri & Begelman (2010). They assumed an effective temperature of 4000 K, while our fiducial model has 8500 K. We compare two different exposure times for HST and JWST, namely $t_{\rm exp} = 10$ h and $t_{\rm exp} = 10^4$ s, respectively. For comparison, the longest exposure in the F160W band of the Hubble Ultra Deep Field '09 captured with the WFC3 camera is ≈ 41 hours[8]. This choice provides a very similar magnitude limit of ≈ 27.8 for both instruments and filters.

We find that hotter quasi-stars are brighter at higher redshift in the considered bands. This is because at $T_{\rm phot} = 4000$ K, the peak of the spectrum is red enough that at high redshift it gets displaced beyond the band limit of the filters. This happens at $z_{\rm source} > 3$ for HST, while at $z_{\rm source} >$ 5-6 for JWST, because the F444W filter extends more in the near-infrared than the F160W one. Quasi-stars with $L_{\rm phot} > 10^{45}$ erg s^{-1} could be in principle detected by both HST and JWST with the considered integration times at $z_{\rm source} \gtrsim 10$. However, they are close to the magnitude limit for HST, while they are well above the same limit for JWST. This suggests that it is fairly unlikely that HST has already observed such a source even within a Ultra-Deep-Field-like exposure, whereas we expect JWST to be able to detect quasi-stars in the luminosity range $10^{44} - 10^{46}$ erg s^{-1} at $z_{\rm source} > 10$. Nonetheless, as Volonteri & Begelman (2010) have shown, there might be some rare events of quasi-stars forming at redshift as low as $z_{\rm source} \sim 4$. Though the bulk of the population is expected to be in place at higher redshift, HST might still observe such an outlier. However, we caution that those numbers represent the most optimistic estimates since we are neglecting the effect of the environment where quasi-stars are expected to live. Indeed, when quasi-stars are harboured within gas-rich environment, part of their radiation might be absorbed and reprocessed in different wavelengths.

3.4 The model of rotating quasi-stars

3.4.1 Quasi-stars as loaded polytropes

As already discussed, the hydrostatic structure of a quasi-star is constituted by a radiation-dominated, convective envelope, surrounded by a thin, radiative layer (see Section 3.3; Begelman, Rossi & Armitage 2008; Ball et al. 2011; Dotan, Rossi & Shaviv 2011). Since the envelope represents the majority of the mass and volume of a quasi-star and convective regions can be described accurately by an adiabatic temperature gradient, a quasi-star can be modelled as a polytropic gas with index $n = 3$. A polytropic gas is characterised by a barotropic equation of state $P(\rho) = P_{\rm c}(\rho/\rho_{\rm c})^{\gamma}$, where $P_{\rm c}$ and $\rho_{\rm c}$ are the central pressure and density, respectively, and the adiabatic index $\gamma = 1 + 1/n = 4/3$ for

---

[8] `https://archive.stsci.edu/prepds/hudf09/`

a radiation-dominated gas. Polytropes are regular solutions of the Lane-Emden equations with inner boundary conditions in the standard dimensionless density and mass variables $\Theta_\mathrm{c} = \Theta(0) = 1$ and $\phi_\mathrm{c} = \phi(0) = 0$. When $n < 5$, they extend up to the dimensionless radius $\xi_\star = r_\star/\alpha$, where $\Theta_\star = \Theta(\xi_\star) = 0$ and α is the standard radial normalisation, and they enclose a total, finite mass $M_\star = 4\pi\rho_\mathrm{c}\alpha^3\phi_\star$ (e.g. Ball, Tout & Żytkow 2012).

Additionally, quasi-stars are characterised by the presence of a central black hole of mass $M_\bullet$. We can model this feature by changing the inner boundary conditions: we assume that within the radius r_0, the enclosed mass is $M(r_0) = M_\bullet$ and that the density and the pressure are normalised to the values ρ_0 and P_0 at r_0, respectively. The radius r_0 is the size of the gravitational sphere of influence of the black hole and is typically of the order of its Bondi radius r_B:

$$r_0 = br_\mathrm{B} = b\frac{GM_\bullet}{2c_{\mathrm{s},0}^2}, \tag{3.45}$$

where $c_{\mathrm{s},0}^2 = \gamma P_0/\rho_0$ and b is a numerical constant of the order of few. In terms of dimensionless quantities, the new boundary conditions at $\xi_0 = r_0/\alpha$ are $\Theta(\xi_0) = \Theta_0 = 1$ and $\phi_0 = \phi(r_0) = M_\bullet/(4\pi\rho_0\alpha^3)$. A polytropic solution with non-zero central mass (i.e. with the latter boundary conditions) is called *loaded polytrope* (Huntley & Saslaw, 1975). Throughout the rest of the paper, we use loaded polytropes to model the internal, hydrostatic structure of a quasi-stars assuming $n = 3$.

We note that ξ_0 and ϕ_0 are not independent, but they are related by:

$$\xi_0 = \frac{3b}{2}\phi_0. \tag{3.46}$$

Therefore, the boundary conditions can be fully determined by choosing a value for ϕ_0. In turn, this is related through the Lane-Emden equation to the total mass of the envelope $\phi_\star$. This relation is shown in Figure 3.8 in terms of the the mass ratio $q \equiv M_\bullet/M_\star = \phi_0/\phi_\star$ as a function of ϕ_0 for different values of b. The mass ratio q has always a maximum at $\phi_0 = \tilde{\phi}_0$. This occurrence has been described in details by Ball, Tout & Żytkow (2012) as a generalisation of a Schönberg-Chandrasekhar-like limit for polytropic envelopes surrounding a central core (Schönberg & Chandrasekhar, 1942). Quasi-stars have typically $q < 10^{-2}$. Solutions on the $\phi_0 > \tilde{\phi}_0$ branch are unphysical because they reach zero mass before zero radius. Acceptable solutions lie on the $\phi_0 < \tilde{\phi}_0$ branch, where the dependency on b becomes very weak. On this branch, we find empirically $\phi_0 \approx 2q$, as shown in Figure 3.8. From this relation, we can build any solution as follows. First, we choose a value of q, typically between $\sim 10^{-4}$ and $\sim 10^{-2}$. This maps to the value of ϕ_0 necessary to set the boundary conditions and specify $\phi_\star$. We can then rescale the dimensionless solution with a specified q to any solution in physical units by specifying the central black hole mass $M_\bullet$ and the pressure P_0. The density ρ_0 can then be obtained as:

$$\rho_0 = \left[\frac{(n+1)^3}{4\pi G^3}\right]^{1/4} \frac{\phi_0^{1/2} P_0^{3/4}}{M_\bullet^{1/2}} \approx 1.2 \times 10^{-5}\ q^{1/2}\ p_{0,8}^{3/4}\ m_\bullet^{-1/2}\ \mathrm{g\ cm^{-3}}, \tag{3.47}$$

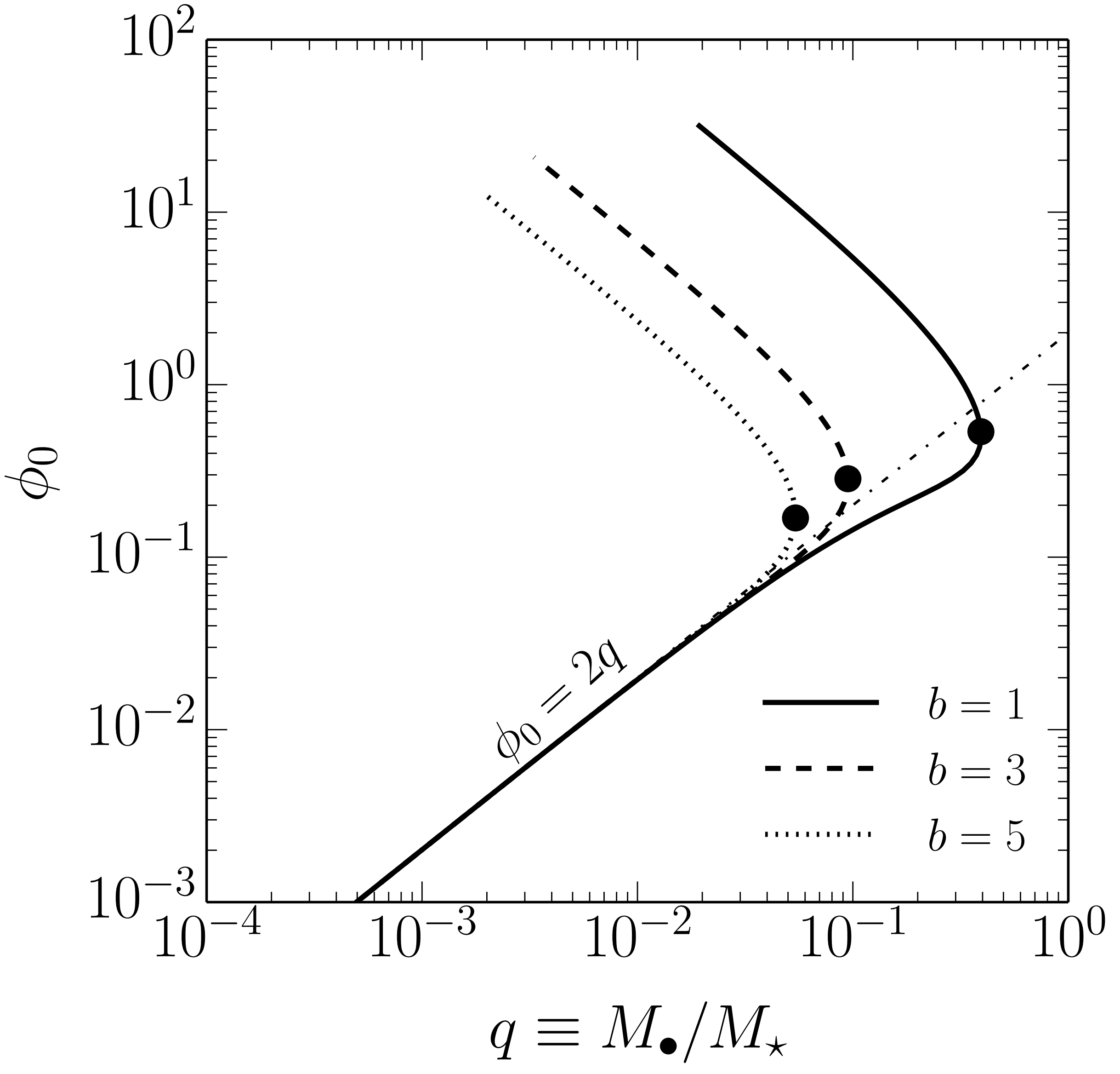

Figure 3.8 Relation between the mass ratio $q = M_\bullet/M_\star = \phi_0/\phi_\star$ and the dimensionless black hole mass ϕ_0. Continuous, dashed and dotted lines show the case $b = 1$, $b = 3$ and $b = 5$, respectively. The black dots show the position of the limiting mass ratio q. The lower branch ($\phi_0 < 0.1$) is the only one that gives physical solutions (see the text) and it almost does not depend on b. The dash-dotted line shows the relation $\phi_0 = 2q$.

where we use $n = 3$, $M_\bullet = m_\bullet$ $\mathrm{M}_\odot$, and $\phi_0 \approx 2q$.

The simplified description of the quasi-star internal structure as a loaded polytrope requires three parameters to be specified, namely the central pressure P_0, the black hole mass $M_\bullet$ and the mass of the envelope $M_\star$ through the ratio q. However, this neglects the energy production mechanism at the centre, which would introduce an additional relation between e.g. P_0 and $M_\star$. Nonetheless, this treatment provides the correct estimates as long as, for each $M_\bullet - q$ pair, P_0 is chosen consistently with detailed equilibrium models such as those presented in Section 3.3. In fact, choosing a consistent value of P_0 given $M_\star$ and $M_\bullet$ is implicitly equivalent to consider the energy transport within the star. Moreover, the convective envelope of equilibrium models of Section 3.3 is formally obtained by solving the equations of a loaded polytope, and since it dominates the mass and volume of the stars, it provides alone a remarkable description of the entire hydrostatic structure.

3.4.2 Differential rotation inside quasi-stars

In a recent series of papers, Balbus and collaborators have developed a theory to describe the convective zone in the Sun (Balbus et al., 2009; Balbus & Weiss, 2010; Balbus & Latter, 2010; Balbus, Latter & Weiss, 2012; Balbus & Schaan, 2012). Their model successfully reproduces the isorotation contours within the solar convective zone and the tachocline from the helioseismology data of the Global Oscillation Network Group (GONG). Here, we review the main features of the model and we then apply it to quasi-stars, mostly following Balbus et al. (2009) and Balbus & Weiss (2010).

We consider an azimuthally rotating, convective gas flow (generically a star) in spherical coordinates (r, θ, ϕ), where r is the radial distance from the centre, θ is the colatitude angle and ϕ is the azimuthal angle. The flow is symmetric with respect to the rotation axis, i.e. the thermodynamic variables characterising it, such as the density ρ, the pressure P and the specific entropy s, do not depend on ϕ, but they generally depend on r and θ. The only velocity component is the azimuthal velocity $v_\phi = r\ \sin\theta\ \Omega(r, \theta)$, where $\Omega(r, \theta)$ is the angular velocity. We neglect any departure from sphericity, implicitly assuming slow rotation. Such a flow in steady state is described by the following Euleur equations (r- and θ-component, respectively, while the azimuthal component is $0 = 0$):

$$\begin{cases} -\dfrac{1}{\rho}\dfrac{\partial P}{\partial r} - \dfrac{\partial \Phi}{\partial r} &= 0, \\ \dfrac{v_\phi^2 \cot\theta}{r} - \dfrac{1}{\rho}\dfrac{\partial P}{\partial \theta} - \dfrac{1}{r}\dfrac{\partial \Phi}{\partial \theta} &= 0, \end{cases} \tag{3.48}$$

where Φ is the gravitational potential. Note that in the radial direction we neglect the (weak) centrifugal force[9]. By taking the ϕ-component of the curl of the Euler equations (equation 3.48), and dropping terms proportional to $\partial P/\partial\theta \ll \partial P/\partial r$, we obtain the thermal wind equation (Kitchatinov & Ruediger, 1995; Thompson et al., 2003; Balbus et al., 2009; Balbus & Weiss, 2010; Balbus, Latter & Weiss, 2012):

$$\frac{\partial \Omega^2}{\partial r} - \frac{\tan\theta}{r}\frac{\partial \Omega^2}{\partial \theta} = \frac{1}{\gamma\, r^2\ \sin\theta\ \cos\theta}\frac{\mathrm{d}\Phi}{\mathrm{d}r}\frac{\partial \sigma}{\partial \theta}, \tag{3.49}$$

---

[9] In fact, this approximation is only necessary to derive equation (3.49) *after* the ϕ-component of the curl of those Euler equations has been taken. For clarity, however, we already drop the centrifugal force at this early step.

where we have introduced the dimensionless entropy function:

$$\sigma = \log \left[\frac{P}{P_0} \left(\frac{\rho}{\rho_0} \right)^{-\gamma} \right], \tag{3.50}$$

which is proportional to (or monotonically dependent on) s. Equation (3.49) neglects the contribution from convective turbulence to the velocity field (it is in fact a time-averaged description of the flow) and it is not valid for highly magnetised stars, but a weak magnetic field can be accommodated (Balbus, 2009).

Let us now introduce the *residual* entropy: the azimuthally averaged entropy profile left, after the radial profile σ_r has been subtracted off:

$$\sigma'(r, \theta) = \sigma - \sigma_r(r). \tag{3.51}$$

Since equation (3.49) depends on σ exclusively through its θ derivative, the differential profile $\Omega(r, \theta)$ could be determined after knowing σ', regardless of σ_r. Convection in a non-rotating star establishes a stable entropy radial profile σ_r, as a result of an equilibrium reached between central stellar heating and heat transport. Convective cells moves on average along the radial direction. If now a small amount of rotation is added, the convective cells will tend, on average, to drift towards surfaces of constant angular rotation. This is because differential rotation tends to confine the flow in sheet of constant Ω. This assumes of course that the rotational surfaces can effectively interact with the convective cells during their lifetime, which is reasonable if they are long lasting structures. In the presence of a relative small degree of rotation, we can therefore argue that σ_r is similar to that established in a non-rotating star, while $\sigma'(r, \theta)$ is a small departure from σ_r, closely connected to the differential rotational profile within the star. Following Balbus et al. (2009), we assume, $\sigma' = f(\Omega^2)$ which implies that surfaces of constant residual entropy coincide with surfaces of constant angular velocity. Though still not unambiguously demonstrated, this conjecture provides remarkable results when used to describe the solar convective zone (Balbus & Latter, 2010; Balbus, Latter & Weiss, 2012; Balbus & Schaan, 2012). In addition, it is also supported, at least qualitatively, by the results of hydrodynamical simulations showing similarity between constant Ω and σ' contours (Miesch, Brun & Toomre 2006; see also figure 2 from Balbus et al. 2009).

With this relation, $\sigma' = f(\Omega^2)$, equation (3.49) can be rewritten as:

$$\frac{\partial \Omega^2}{\partial r} - \left(\frac{\tan\theta}{r} + \frac{f'}{\gamma\, r^2 \, \sin\theta \, \cos\theta} \frac{\mathrm{d}\Phi}{\mathrm{d}r} \right) \frac{\partial \Omega^2}{\partial \theta} = 0, \tag{3.52}$$

where $f' = \mathrm{d}\sigma'/\mathrm{d}\Omega^2$. The above equation has the form $\mathbf{u} \cdot \nabla\Omega^2 = 0$, where $\mathbf{u}$ is the vector tangential to the surfaces of constant Ω (i.e. it is their "velocity" vector). Such surfaces $\boldsymbol{\zeta} = (r, \theta(r))$ can be obtained by integrating the ordinary differential equation $\dot{\boldsymbol{\zeta}} = \mathbf{u}$ (where $\dot{}$ indicates the derivative with respect to any dummy parameter). More practically, one divides the polar and radial component of that vectorial equation and obtains the following single equation:

$$\frac{\mathrm{d}\theta(r)}{\mathrm{d}r} = \frac{\tan\theta}{r} + \frac{f'}{\gamma\, r^2\, \sin\theta\, \cos\theta}\frac{\mathrm{d}\Phi(r)}{\mathrm{d}r}. \tag{3.53}$$

If we recall that f' depends only on Ω^2 and that equation (3.53) describes surfaces of constant Ω^2, we can finally integrate the above equation considering f' as constant:

$$r^2\sin^2\theta = A - \frac{2f'}{\gamma}\Phi(r), \tag{3.54}$$

where A is an integration constant. These iso-Ω^2 surfaces are the characteristics of equation (3.52). Note that on each surface, f' can assume a different constant value.

We can determine the constant A by specifying a starting position for each characteristic. We take the position $(R_\star, \theta_\star)$ at the surface of a spherical star with radius $R_\star$ and we obtain:

$$r^2\sin^2\theta = R_\star^2\sin^2\theta_\star - \frac{2f'}{\gamma}(\Phi(r) - \Phi(R_\star)), \tag{3.55}$$

where now $f' = f'(\Omega(R_\star, \theta_\star))$ has to be specified and the internal structure of the star influences the result through Φ. The curves described by equation (3.55) are constant Ω contours, therefore they can be used to reconstruct the 2-dimensional $\Omega(r,\theta)$ by assigning a value of Ω at a given radius. Specifically, we will supply $\Omega_\star(\theta_\star)$ at $R_\star$. We can then isolate $\theta_\star(r,\theta)$ from equation (3.55) and obtain $\Omega(r,\theta) = \Omega_\star(\theta_\star(r,\theta))$.

We can now use this method to explicitly calculate the internal differential rotation of quasi-stars, once we specify their internal structure. Since we describe quasi-stars as loaded polytropes (see Section 3.4.1), we can integrate the equation of hydrostatic equilibrium between r and $R_\star$ for a polytropic equation of state and obtain:

$$\Phi(r) - \Phi(R_\star) = -\frac{c_\mathrm{s}^2}{\gamma - 1} = -3\ c_{\mathrm{s},0}^2\ \Theta_3(r;q), \tag{3.56}$$

where $c_{\mathrm{s},0}^2 = \gamma P_0/\rho_0$ is the sound speed at r_0, $\Theta_3(r;q)$ is the loaded polytrope solution for $n = 3$ and a given q, and $\gamma = 4/3$. Substituting equation (3.56) into equation (3.55), we finally get:

$$\sin^2\theta_\star = \left(\frac{r}{R_\star}\right)^2 \sin^2\theta - \beta\ \Theta_3(r;q), \tag{3.57}$$

where we define:

$$\beta = \frac{9c_{\mathrm{s},0}^2 f'}{2R_\star^2}. \tag{3.58}$$

There is still a quantity that has remained general in our treatment, namely $f'(\Omega^2)$. Unfortunately, we do not know a priori its functional form, and only hydrodynamical simulations of global 3D convection could clarify this point. However, to avoid unnecessary mathematical complication at this stage, we assume the simplest functional form, i.e. a global constant for f'. We can now use simple reasonable arguments to constrain the constant parameter β, that directly depends on f' through equation (3.58).

First, we expect $\beta < 0$ (i.e. $f' < 0$, like in the convective envelope of the Sun), since this implies slower rotating poles with respect to the equatorial regions. This configuration may naturally comes about when quasi-stars are fed by protogalactic discs near the equator, i.e. angular momentum is injected by the infalling material near the equator and has to be redistributed from there to the poles. Finally, we can also estimate the value of $|\beta|$ by recalling that σ' is a small perturbation on the otherwise spherically symmetric entropy profile σ_r which arises when the star rotates:

$$\sigma' \sim \frac{T_{\rm rot}}{U} \sim \frac{R_\star^2 \Omega^2}{c_{\rm s,0}^2}, \tag{3.59}$$

where $T_{\rm rot} \sim M_\star R_\star^2 \Omega^2$ and $U \sim M_\star c_{\rm s,0}^2$ are the rotational kinetic energy and the gaseous internal energy of the star, respectively. Therefore, $f' \sim \sigma'/\Omega^2 \sim R_\star^2/c_{\rm s,0}^2$ implies that $|\beta| \sim 9/2 \sim$ a few. Although this simple line of reasoning does not prove that f' should be constant, it provides a gross estimate of the value of $|\beta|$ *if* f' is assumed to be constant. However, we show in Section 3.4.3 that the exact value of $|\beta|$ has a weak impact on our final conclusions and we discuss the limitations of our approach in Section 3.4.6.

3.4.3 Angular velocity structure of quasi-stars

To explicitly calculate the differential rotation within a quasi-star, we need to specify the boundary conditions of the problem, i.e. the differential rotation at the surface $\Omega_\star(\theta_\star)$. We use a simple parametrisation of the form:

$$\Omega_\star(\theta_\star) = \omega_\star(1 + \delta\omega \sin^2\theta_\star), \tag{3.60}$$

where $\omega_\star$ is the polar rotation, limited by the Keplerian velocity of the star $\Omega_{\rm K,\star}$, and $\delta\omega$ is the relative, fractional excess of rotation at the equator, with the limit $0 < \delta\omega < \epsilon^{-1} - 1$, where $\epsilon = \omega_\star/\Omega_{\rm K,\star} < 1$. This parametrisation of the differential rotation has been used to describe the Sun as well as other stars, with typical values $\delta\omega \sim 0.1$ (e.g. Balbus et al. 2009; Reinhold, Reiners & Basri 2013).

Figure 3.9 shows the angular velocity profiles and maps for a reference quasi-star with $q = 10^{-3}$ (e.g. a massive quasi-star with $M_\bullet = 10^3$ M$_\odot$ and $M_\star = 10^6$ M$_\odot$), $\epsilon = 0.5$ and $\delta\omega = 0.2$ (i.e. rotating at 0.6 $\Omega_{\rm K,\star}$ at the equator, with a differential velocity of 20% between the equator and the poles), and we vary the value of $|\beta|$ between 1 and 16. The upper-left panel (sub-panel (a)) shows the radial profile of the θ-averaged angular velocity $\langle\Omega\rangle$ (normalised by $\Omega_{\rm K,\star}$), highlighting its behaviour at small radii. Initially, $\langle\Omega\rangle$ grows from the surface of the star inward for most of the stellar volume till $\sim 0.1 R_\star$. This growth is accentuated for larger values of $|\beta|$. Within $0.1R_\star$, $\langle\Omega\rangle$ remains almost constant, assuming a solid-body-like rotation law and following the central density of the

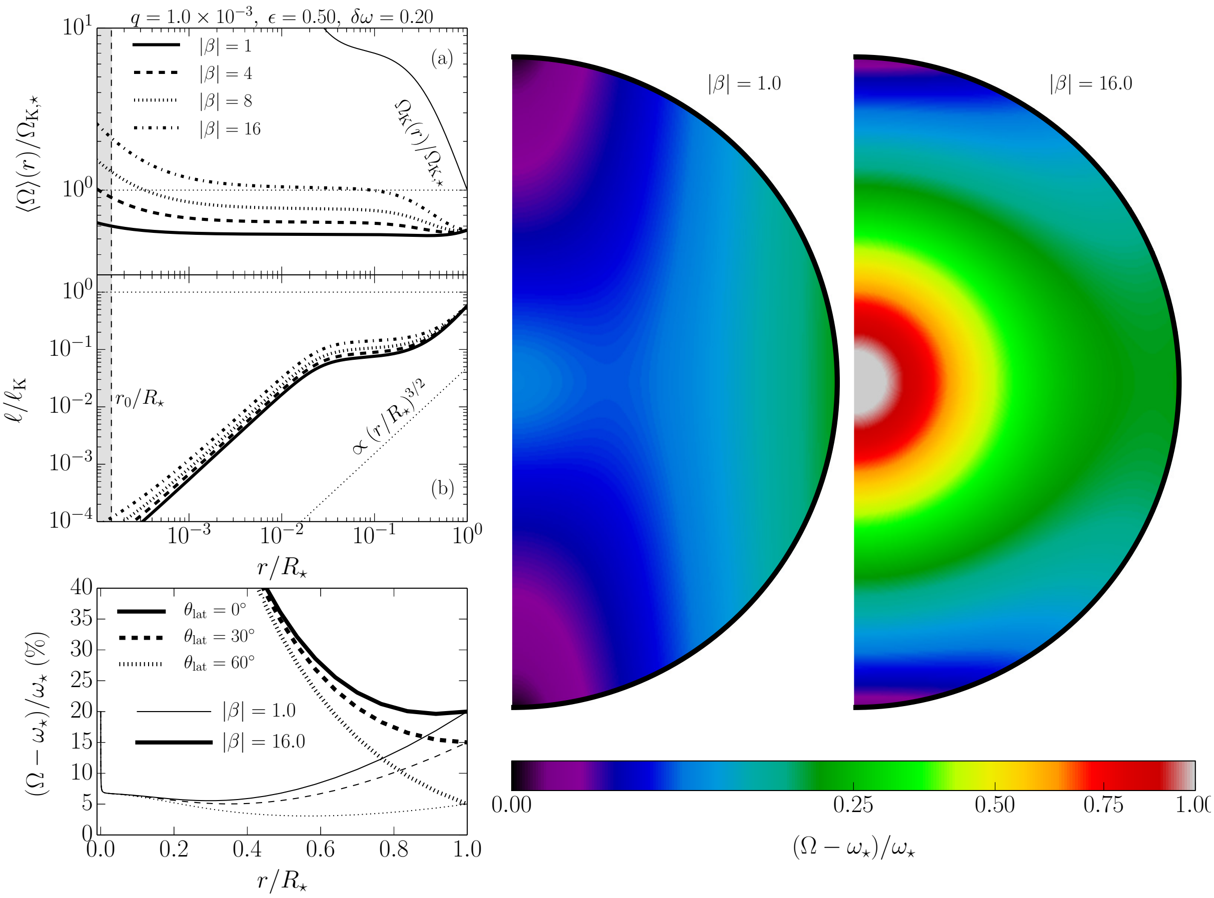

Figure 3.9 Angular velocity for a quasi-star model characterised by $q = 10^{-3}$, $\epsilon = 0.5$ and $\delta\omega = 0.2$. Top-left panel: radial profile of the θ-averaged angular velocity (sub-panel (a)), normalised to the Keplerian angular velocity $\Omega_\star$, and of the ratio ℓ/ℓ_{K} of the θ-averaged specific angular momentum over the Keplerian one for the same mass distribution (sub-panel (b)). Solid, dashed, dotted, and dash-dotted lines correspond to $|\beta| = 1, 4, 8, 16$, respectively. The solid thin line in sub-panel (a) shows the Keplerian angular velocity. Bottom-left panel: percentage excess of rotation relative to the polar angular velocity as a function of radius at constant latitude θ_{lat}. Solid, dashed, and dotted lines correspond to $\theta_{\mathrm{lat}} = 0°$ (equator), $30°$, and $60°$, respectively, while thin and thick lines refer to $|\beta| = 1$ and $|\beta| = 16$. Right panel: 2D maps of the internal relative excess of angular velocity relative to the polar rotation for $|\beta| = 1$ and $|\beta| = 16$. Note that the surface rotation of both maps is the same.

gas that also starts to flatten in a central core. However, the angular velocity deviates from constant within 0.005 $R_\star$ (or $\sim 20\ r_0$) because of the presence of the central black hole, steepening at smaller radii. The typical trend is $\propto r^{-\zeta}$, with $\zeta \sim 0.2 - 0.6$, increasing with $|\beta|$. We compare $\langle\Omega\rangle$ with the Keplerian angular velocity $\Omega_{\mathrm{K}}(r) = \sqrt{GM(r)/r^3}$ associated to the same mass distribution. Outside $\sim 0.2\ R_\star$, Ω_{K} grows inward as $r^{-3/2}$ (since most of the mass of the envelope is contained in the central core), faster than $\langle\Omega\rangle$. Then, it slightly flattens, but it suddenly starts to grow again as $r^{-3/2}$ due to the presence of the central black hole that dominates the enclosed mass ϕ out to $\sim 0.02 R_\star$, resulting in $\Omega_{\mathrm{K}} \gg \langle\Omega\rangle$ at r_0.

Although convection can induce solid body rotation, this is not achieved in the entire envelope, but only in the central part. This is shown in the lower-left panel of Figure 3.9, where we plot the radial profiles of Ω (shown as the percentage excess of rotation compared to the surface angular velocity at the poles $\omega_\star$) at different latitudes $\theta_{\rm lat}$ ($\theta_{\rm lat} = 0^\circ$ means the equator). Most of the stellar volume is differentially rotating at different latitudes, as shown by the two extreme examples $|\beta| = 1$ and $|\beta| = 16$. Those are representative of the two limiting cases: when $|\beta| \to 0$, the angular velocity becomes constant on cylinders. This can be seen in the region close to the surface around the equator of the map corresponding to $|\beta| = 1$. At constant latitude, Ω decreases as r goes from the surface to $\sim 0.3 - 0.4R_\star$, when it starts to mildly grow inward and it becomes nearly constant within $\sim 0.1R_\star$; then, it steepens again close to the central black hole. On the other hand, when $|\beta| \gg 1$, the angular velocity tends to be "shellular", i.e. it mostly follows the isobars and Ω varies with r only. That can be seen in the example map for $|\beta| = 16$ within $\sim 0.5R_\star$, while in the outermost parts of the star the angular velocity maintains a net θ-dependency and it quickly grows as r decreases.

The ratio $\ell/\ell_{\rm K}$ between the specific angular momentum ℓ and the Keplerian angular momentum for the same mass distribution closely relates to $\langle\Omega\rangle$ and $\Omega_{\rm K}$, as shown in the top-left panel (sub-panel (b)). Outside $\sim 0.2R_\star$, $\ell/\ell_{\rm K}$ first decreases inward, then flattens till $\sim 0.02R_\star$, and finally starts to decrease again, roughly as $\propto r^{3/2}$. Within $0.005R_\star$, the ratio deviates from the previous scaling and decreases more slowly, as a consequence of the steepening of $\langle\Omega\rangle$ close to the central black hole (Figure 3.9). Close to r_0, $\ell/\ell_{\rm K}$ typically assumes values $\sim 10^{-4}$, with mild variations within a factor $\gtrsim 2$ when $|\beta|$ is changed between 1 and 16, suggesting that the exact value of $|\beta|$ has a minor impact on the ratio $\ell/\ell_{\rm K}$ at about r_0.

We have also tested the sensitivity of this behaviour, and in particular of $\ell/\ell_{\rm K}$ around r_0, on the structural parameters of the star, namely q, ϵ and $\delta\omega$. Quasi-stars with proportionally larger black holes at the centre (i.e. with larger q) tend to have larger $\ell/\ell_{\rm K}$ close to r_0, though this ratio remains confined within a few 10^{-3} for $q \sim 0.1$. $\ell/\ell_{\rm K}$ departs more from $\propto r^{3/2}$ for larger q, becoming shallower close to r_0. When the quasi-star rotates proportionally faster at the equator than at the pole (i.e. when ϵ decreases and $\delta\omega$ increases) the spread between the rotation laws with different $|\beta|$ increases, but this does not change their shape and the typical $\ell/\ell_{\rm K}$ close to r_0. As an (extreme) example, we consider a quasi-star that almost reaches the break-up velocity at the equatorial surface (with $\Omega_{\rm eq}/\Omega_{{\rm K},\star} = 0.8$), but rotates slowly at the poles (with $\epsilon = 0.1$), the curve with $|\beta| = 16$ lies almost a factor of 10 above the curve with $|\beta| = 2$. Still, also in this example, $\ell/\ell_{\rm K} <$ a few 10^{-3} at r_0. Of course, such a level of differential rotation, with the equatorial velocity close to the break-up velocity, might induce an appreciable departure from sphericity that our model does not take into account and therefore we can not exclude additional effects in these regime, though the central regions might not be too affected, always retaining a larger degree of sphericity.

We therefore conclude that our calculations suggest that, regardless of the strict values of the parameters assumed, the typical specific angular momentum where the gravity of the black hole starts to dominate (i.e. around r_0) is a fraction not higher than a few $\times 10^{-3}$ of the local Keplerian angular momentum.

3.4.4 Possible implications on the fate of supermassive stars

So far, we have investigate the role of rotation within quasi-stars. Our treatment of rotation can be used to assess the coupling between the inner accretion region and the massive envelope and therefore to speculate on the fate that a supermassive star would undergo, being a quasi-star the possible next evolutionary stage. When rotation is not included, the evolution of a quasi-star is discussed in Section 3.3 and summarised by the results of Figure 3.5 and 3.6. However, our results of Section 3.4.3 might affect this picture.

In our calculations, we neglect any general relativistic effect. It is then worth comparing the Schwarzschild radius $r_{\rm s}$ of the central black hole, that sets the size of the black hole horizon, with the envelope's inner radius r_0, where our calculation stops:

$$\begin{aligned} \frac{r_{\rm s}}{r_0} &= \frac{4c_{\rm s,0}^2}{bc^2} = \frac{8\,(\pi G^3)^{1/4}}{3\;c^2}\;\frac{M_\bullet^{1/2}\;P_0^{1/4}}{b\;\phi_0^{1/2}} \approx \\ &\approx 5.2\times10^{-8}\;b^{-1}\;q^{-1/2}\;m_\bullet^{1/2}\;p_{0,8}^{1/4}, \end{aligned} \tag{3.61}$$

where we used $\phi_0 \approx 2q$. Inserting consistent mass-pressure values from the models of Section 3.3 we typically find that r_0 is between a few to several thousand Schwarzschild radii of the black hole. For example, when we consider the Growth Region and take (i) a relatively small quasi-star, ($M_\bullet = 100$ M$_\odot$, $M_\star = 2\times10^5$ M$_\odot$ and $P_0 = 6.3\times10^{10}$ erg cm^{-3}); and (ii) a massive quasi-star with a relatively more massive black hole, ($M_\bullet = 10^4$ M$_\odot$, $M_\star = 10^7$ M$_\odot$ and $P_0 = 3.4\times10^{10}$ erg cm^{-3}), we find $r_{\rm s}/r_0 \approx 1.2\times10^{-4}$ and $r_{\rm s}/r_0 \approx 7.1\times10^{-4}$, respectively. These estimates support our choice of neglecting any general relativistic effect and we can therefore safely use our results at r_0 to put boundary conditions to the central accretion flow.

Although a detailed modelling of the central accretion flow is beyond the purpose of this work, we can still gain insight into its formation and some possible features from simple inferences from our results. The results of Section 3.4.3 suggest that the specific angular momentum at r_0 is $\ell_0 = \eta\;\ell_{\rm K}(r_0)$, where η is $\sim 10^{-3}-10^{-4}$. By assuming the conservation of angular momentum, we can calculate the circularisation radius $r_{\rm circ}$ around the central black hole, i.e. the radius at which ℓ_0 corresponds to a circular orbit:

$$\frac{r_{\rm circ}}{r_0} = \eta^2, \tag{3.62}$$

where we assume that $\Omega \simeq \Omega_{\mathrm{K}}$ below r_0, as the black hole's gravity dominates. This radius tells us the scale below which some sort of accretion disc may eventually form. That requires $r_{\mathrm{circ}} > r_{\mathrm{isco}} \approx r_{\mathrm{s}}$, where r_{isco} is the radius of the innermost stable circular orbit, which is a few times r_{s}, depending on the black hole spin. We can combine equation (3.61) and (3.62) to determine a condition on η,

$$\begin{aligned} \eta \; > \; \eta_{\mathrm{crit}} = & \left[\frac{8\,(\pi G^3)^{1/4}}{3\,c^2}\right]^{1/2} \frac{M_\bullet^{1/4} P_0^{1/8}}{b^{1/2}\phi_0^{1/4}} \approx \\ \approx \; & 2.3\times10^{-4}\, b^{-1/2}\, q^{-1/4}\, m_\bullet^{1/4}\, p_{0,8}^{1/8}. \end{aligned} \tag{3.63}$$

When $\eta > \eta_{\mathrm{crit}}$, the gas circularisation is such that $r_{\mathrm{circ}} > r_{\mathrm{isco}}$ and an accretion disc can form at the centre of a quasi-star. As an example, we calculate η_{crit} for the same quasi-star models used above, and find that $\eta_{\mathrm{crit}} \approx 1.1\times10^{-2}$ and $\eta_{\mathrm{crit}} \approx 2.7\times10^{-2}$, respectively. Those numbers are also representative of the whole Growth Region, as they weakly depend on q, $M_\bullet$ and P_0 (see equation 3.63). Interestingly, $\eta_{\mathrm{crit}} \gtrsim 10^{-2}$ is comfortably larger than the indicative upper limit on $\eta \sim$ a few $\times\, 10^{-3}$ that we estimate in Section 3.4.3, leading to the conclusion that typical quasi-stars in the Growth Region might *not* be able to develop an accretion disc at their centre.

To better assess this point, we exploit the models used in Section 3.3 to thoroughly explore the parameter space, as shown in Figure 3.10 . Specifically, we take the values of $M_\bullet$, $M_\star$ (hence q), and P_0 (calculated at $r_0 = 5r_{\mathrm{B}}$) and we use them to calculate η and η_{crit} across the $M_\bullet - M_\star$ plane. As discussed in Section 3.4.3, the value of η depends on some parameters, namely $(|\beta|, \epsilon, \delta\omega)$. We tested several configurations: (i) a "fiducial" model with $(4.5, 0.5, 0.2)$, (ii) a model with a higher value for $|\beta|$, $(12, 0.5, 0.2)$, (iii) a rapidly and differentially rotating quasi-star with $(4.5, 0.1, 7.0)$, and (iv) a slowly rotating quasi-star with $(4.5, 0.1, 0.2)$. In all cases, we find that for each $M_\bullet$ there is an upper limit on the mass of the envelope above which the condition $\eta > \eta_{\mathrm{crit}}$ is not satisfied, i.e. a rotationally-supported accretion flow cannot form. These limits are shown as thick solid lines with downward pointing arrows in Figure 3.10. We note that the higher limits correspond to faster surface rotation and larger values of $|\beta|$. Fitting the upper limit lines, we obtain that a disc *cannot* form for:

$$M_\star \gtrsim 0.9 - 1.3\times10^5\ \mathrm{M}_\odot \left(\frac{M_\bullet}{100\ \mathrm{M}_\odot}\right)^{0.82}, \tag{3.64}$$

where the ≈ 0.7 uncertainty factor accounts for differences due to the parameters described above. It is very interesting to note that most of the allowed region coincides with the Evaporation Strip (where the black hole has no time to grow), while the Growth Region (where the central black hole could quickly grow to large masses) is almost entirely excluded.

Our conclusions might affect the evolution of quasi-stars. An accretion disc is required as it provides an efficient source of luminosity to sustain the envelope through transport of

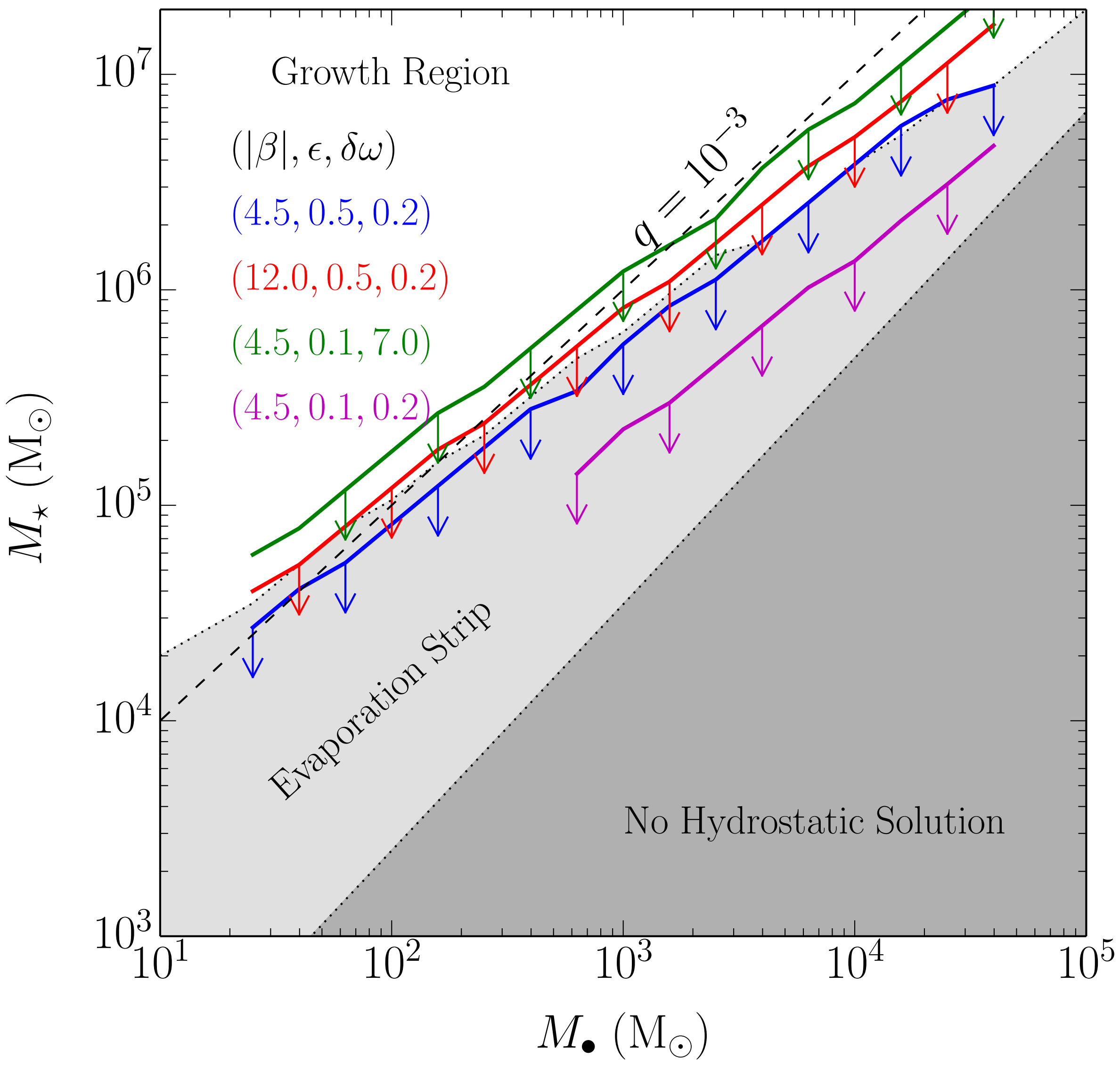

Figure 3.10 $M_\bullet - M_\star$ plane for quasi-stars, divided in three regions: the growth region (white), the evaporation strip (light grey), and the no-hydrostatic solution region (dark grey). Blue, red, green and magenta continuous lines show the upper limits on $M_\star$ as function of $M_\bullet$ for different choices of parameters, namely $(|\beta|, \epsilon, \delta\omega) = (4.5, 0.5, 0.2)$, $(12, 0.5, 0.2)$, $(4.5, 0.1, 7)$, and $(4.5, 0.1, 0.2)$, respectively. Above this limits, which are computed from the models of Section 3.3, the condition $\eta > \eta_{\rm crit}$ is not satisfied. For reference, the black dashed line corresponds to $q = 10^{-3}$.

angular momentum and the extraction of gravitational potential energy (e.g. via magneto-rotational instability; Balbus & Hawley 1991). If a rotationally-supported disc cannot form, we would be in the presence of an optically-thick, quasi-radial flow, that would nearly follow a Bondi-like accretion flow, if well within the trapping radius (Begelman, 1978, 1979). In this case, within the Bondi radius, the gas becomes supersonic and almost free-falling, converting most of the gravitational potential energy into kinetic energy and little into internal energy that could be eventually radiated or convectively transported outward. Since the black hole has no surface, this kinetic energy cannot be dissipated and is advected into the black hole. Even assuming a dissipation mechanisms within the flow, most of the radiation produced would be dragged inward and swallowed by the black hole (Begelman, 1979; Alexander & Natarajan, 2014). Therefore, a consistent model for a quasi-star seems not to exist in the Growth Region of the parameter space of Figure 3.10.

Stepping onto more speculative grounds, we may foreseen a possible fate for supermassive black holes that might form in such conditions. According to recent numerical and analytical calculations, when a very massive star ($\gtrsim 10^5$ $\mathrm{M}_\odot$) forms as consequence of a high accretion rate of gas ($\gtrsim 1$ $\mathrm{M}_\odot$ yr^{-1}), its inner core eventually collapses, presumably into a small (~ 100 $\mathrm{M}_\odot$) black hole (Begelman, 2010; Hosokawa et al., 2013). At this point, however, our results suggest that the surrounding mass might start to be radially accreted, unimpeded by the black hole energy feedback. Since there is no maximum limit for the accretion rate onto a black hole (there is only a limit in luminosity), this may lead to a phase of super-exponential accretion (Alexander & Natarajan, 2014). The process will stop when/if the angular momentum in the accretion flow increases outward so that the circularisation radius increases faster than the black hole's r_{isco}. The outcome clearly depends on the exact hydrodynamics of the flow, but direct formation of massive seeds $\gtrsim 10^5$ $\mathrm{M}_\odot$ might be in principle possible. In this scenario, limiting factors for the black hole seed mass would be linked to the galaxy ability to funnel and accumulate pristine gas in its centre: low cosmological gaseous inflow rate, non efficient angular momentum redistribution and copious star formation (Latif et al., 2013; Choi, Shlosman & Begelman, 2013, 2015).

Small black holes born in the Evaporation Strip might face a different fate. There, an accretion disc can still form, but the available angular momentum is usually low, such that dissipation should occur very close to the black hole. One may therefore speculate that a quasi-spherical, geometrically-thick, radiation-dominated accretion disc, such as a "ZEBRA" (ZEro-BeRnoulli Accretion flow; Coughlin & Begelman 2014) can form. Therefore, we have tried to smoothly join the ZEBRA with the envelopes of our models at the inner radius r_0. First, we note that $\eta^2 = \ell_0^2/\ell_{\mathrm{K}}^2(r_0)$ corresponds to the normalisation of the specific angular momentum of the gas "a" (see equation (10) in Coughlin & Begelman 2014). Given $\gamma = 4/3$, this only depends on the radial slope n of the mass flow within the ZEBRA (i.e. $\dot{M} \propto r^n$), which is the main structural parameter of the model. Since the ZEBRA should form in the central region of the envelope within r_0, we assume that its external radius $\mathcal{R} = r_0$. Finally, we normalise the density by requiring that the luminosity transported by convection outward through the ZEBRA envelope, i.e. $L_{\mathrm{adv}} \approx 4\pi\mathcal{R}^2 P(\mathcal{R}) c_{\mathrm{s}}(\mathcal{R}) = 4\pi\mathcal{R}^2 P_0 c_{\mathrm{s},0}$, is equal to the central luminosity $L_\bullet$ required to self-consistently sustain the envelope. Unfortunately, we find no consistent solutions where the accretion disc is less massive than the black hole, as it is envisaged in the original model. We would therefore need an extension of this model to self gravitating disc, to assess its viability in our case. Another possibility is to relax the requirement that $L_{\rm adv} = L_\bullet$ and speculate instead that $L_\bullet$ is provided by partially tapping the energy funnelled into a powerful jet, whose presence is foreseen in the ZEBRA model. However, the jet is likely going to pierce the envelope, behaving as an outlet of energy, and therefore how enough energy could be transferred in a gentle, uniform way to the envelope is unclear, though possible in principle.

Nonetheless, even if it would be possible to inject within the quasi-star the required luminosity at/above the Eddington limit for the whole mass, the evaporation of the envelope would anyway prevent substantial accretion to occur. Therefore, there might be two populations of supermassive black hole seeds from direct collapse via supermassive stars: one extremely massive, say $> 10^{4-5}$ M$_\odot$ in massive haloes $\gtrsim 10^{8-9}$ M$_\odot$, and one extremely light ~ 100 M$_\odot$ in more common haloes at the epoch of formation ($z \sim 15$). This possibility represents also a “smooth” transition between scenarios of light-seed formation based on PopIII stars and massive-seed formation based on direct collapse.

3.4.5 Direct collapse haloes

We now attempt to identify the haloes that could host a supermassive star $M_\star \gtrsim 10^5$ M$_\odot$. First, we require that $\dot{M} \gtrsim 0.1$ M$_\odot$ yr^{-1}, as needed to assemble $M_\star$ within $\sim 1-2$ Myr, i.e. the lifetime of a supermassive star (Begelman, 2010; Hosokawa et al., 2013). The supermassive star may accrete either gas transported through the protogalactic disc or from cosmological inflows onto the halo, proceeding all the way down to the centre as cold flows (e.g. Di Matteo et al. 2012). The latter case can be translated in a lower limit on the redshift z at which a halo $M_{\rm h}$ can accrete at $\dot{M} > 0.1$ M$_\odot$ yr^{-1} through the relation[10] (Dekel, Sari & Ceverino, 2009):

$$\dot{M} \approx 75 \left(\frac{f_{\rm b}}{0.16}\right) \left(\frac{1+z}{3}\right)^{2.25} \left(\frac{M_{\rm h}}{10^{12}\ {\rm M}_\odot}\right)^{1.15} {\rm M}_\odot\ {\rm yr}^{-1}, \tag{3.65}$$

where $f_{\rm b}$ is the cosmic baryon fraction. We follow Schneider (2015) to calculate the collapse redshift of the halo, i.e. the redshift $z_{\rm col}$ at which a fraction F of the mass $M_{\rm h}$ at redshift $\tilde{z}$ is assembled, by solving the following equation for $z_{\rm col}$:

$$\frac{1}{D(z_{\rm col})} = \frac{1}{D(\tilde{z})} + \sqrt{\frac{\pi}{2}} \frac{\sqrt{\sigma^2(FM_{\rm h}) - \sigma^2(M_{\rm h})}}{\delta_{\rm c,0}}, \tag{3.66}$$

where $D(z)$ is the linear growth factor ($D(0) = 1$), $\delta_{\rm c,0} = 1.686$, and $\sigma^2(M)$ is the present-day variance of the matter density field (i.e. the integral of the linear matter power

---

[10] The usage of equation (3.65) assumes that the gas accretion rate onto the halo is comparable to that onto the forming supermassive star. However, we caution that this approach, in order to keep the calculations simple, neglects the possibility that the supermassive star is at the centre of a protogalaxy.

spectrum over the wavenumber k) at mass scale M (for additional details, see Schneider 2015). Assuming $F = 0.05$, Figure 3.11 shows $z_{\rm col}$ as a function of $M_{\rm h}$ for two values of $\dot{M}$. We adopt the cosmological parameters from the latest Planck results (Planck Collaboration et al., 2015) and we find differences within a factor 2 when we vary F from 0.05 to 0.5.

As a second constraint, we require a metallicity below $\log(Z_{\rm cr}/Z_{\odot}) = -3.8$, where $Z_{\odot}$ is the solar metallicity and the critical value roughly corresponds to the transition from PopIII to second population stars (Valiante et al., 2016). We impose this condition by using the stellar mass-metallicity relation as a function of time determined by Savaglio et al. (2005), and then connecting the stellar mass to $M_{\rm h}$ through the halo mass-stellar mass relation from Moster, Naab & White (2013). After computing the redshift of collapse, we obtain a lower limit $z_{\rm col}(M_{\rm h})$ for haloes $M_{\rm h}$ that have $Z < Z_{\rm cr}$ by the end of the collapse. Finally, the supermassive star cannot be larger than a fraction f of the baryonic mass of the halo, namely $M_{\rm h} > M_{\star}/(f_{\rm b} f)$, where $f \sim 0.01$.

The red shaded region in Figure 3.11 shows where all these conditions are satisfied in the $M_{\rm h} - z_{\rm col}$ plane. We also compare this region with those occupied by haloes with virial temperature $T_{\rm vir} < 10^4$ K and $T_{\rm vir} < 3 \times 10^4$. The virial temperature is calculated as $T_{\rm vir} \approx (G M_{\rm h} H \sqrt{\Delta/54})^{2/3} \mu m_{\rm p}/k_{\rm B}$, where $k_{\rm B}$ is the Boltzmann constant, $m_{\rm p}$ is the proton mass, $\mu \approx 0.59$ is the mean molecular weight for ionised hydrogen, $H(z)$ is the Hubble parameter, and $\Delta(z)$ is the z-dependent virial over-density (Bryan & Norman, 1998).

The latest haloes that might be able to host a supermassive star $M_{\star} > 10^5$ M$_{\odot}$ collapse at $z_{\rm col} \sim 6.5$ and have masses $M_{\rm h} \sim 2 - 3 \times 10^9$ M$_{\odot}$. Those objects represents $\sim 2\sigma$ peaks in the matter density distribution and have typical comoving number densities $\sim 0.6 - 0.9$ cMpc^{-3} dex^{-1}. Supermassive stars can also form within both heavier and lighter haloes virtually at any redshift $z_{\rm col} > 10$, when they are able to sustain the inflow and the gas is still pristine enough. However, beyond $z_{\rm col} \sim 20$, the candidate hosts of supermassive stars more massive than 10^5 M$_{\odot}$ becomes extremely rare, representing more than 5σ over-density fluctuations of the matter density field. Therefore, we can grossly identify the hosts of supermassive stars possibly leading to the formation of $\sim 10^{4-5}$ M$_{\odot}$ back hole seeds as dark mater haloes with masses about $\sim 10^9$ M$_{\odot}$, collapsing between $z \sim 20$ and $z \sim 10$, in agreement with previous results (e.g. Begelman, Volonteri & Rees 2006; Volonteri & Begelman 2010; Valiante et al. 2016). However, we note that our approach (i) requires to extrapolate the used relations to relatively high z, and (ii) it does not account for environmental effects (e.g. the proximity of a massive halo producing H_2-dissociating Lyman-Werner photons), therefore the limits above should be taken as approximated.

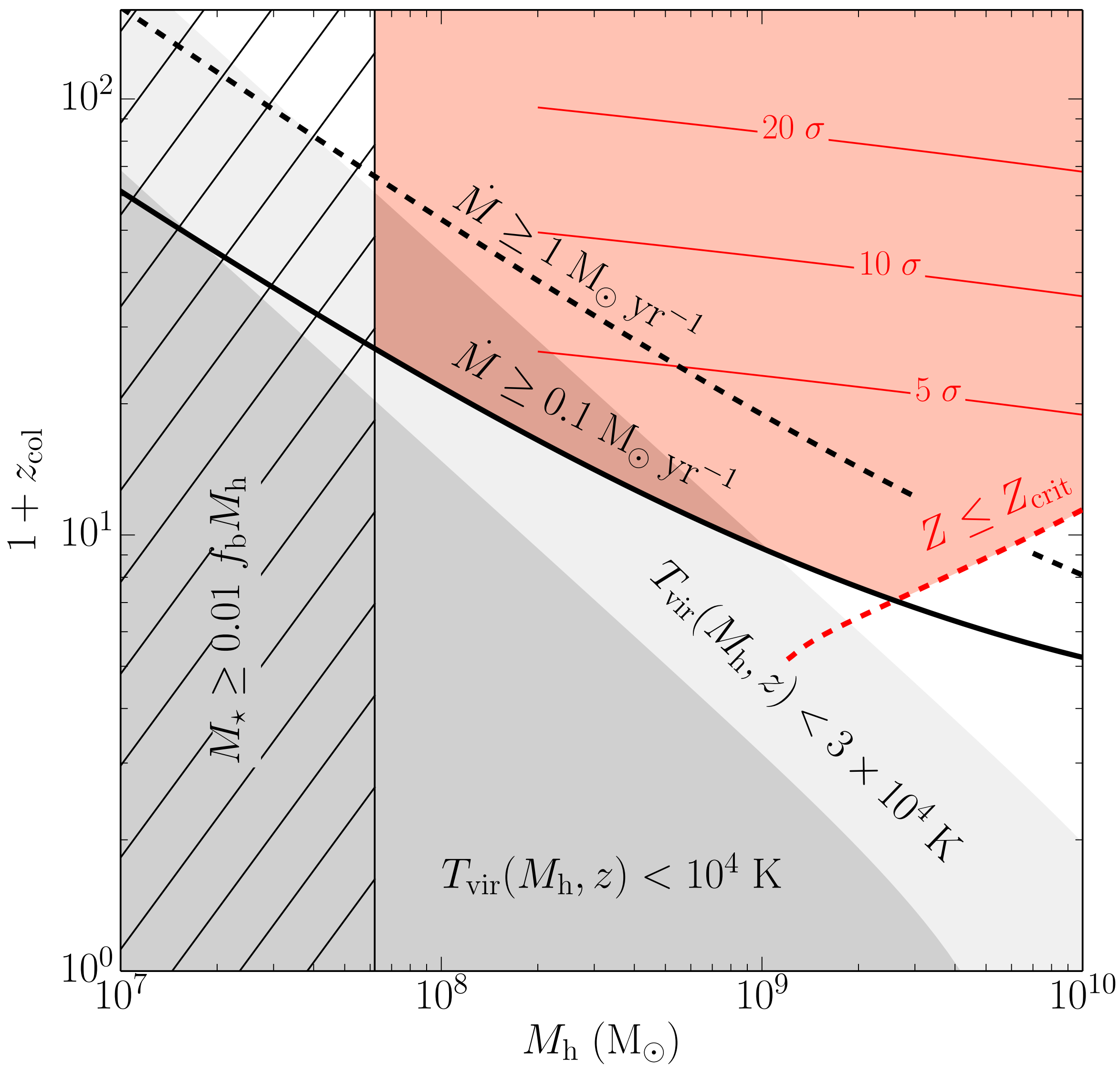

Figure 3.11 Redshift of collapse $z_{\rm col}$ (defined by $F = 0.05$) as a function of the halo mass $M_{\rm h}$. The continuous and dashed lines refer to $\dot{M} = 0.1$ and 1 M$_\odot$ yr^{-1}, respectively. The dark and light grey regions mark where the halo virial temperature $T_{\rm vir} < 10^4$ K and $T_{\rm vir} < 3 \times 10^4$ K, respectively. The hatched region marks the halo masses for which a supermassive star $M_\star = 10^5$ M$_\odot$ represents more than 1% of the baryonic mass $f_{\rm b} M_{\rm h}$. The red dashed line marks the redshift threshold above which a halo $M_{\rm h}$ has $Z < Z_{\rm cr}$ after collapsing. The red shaded region shows where supermassive stars $> M_\star$ could form in the $M_{\rm h} - z_{\rm col}$ plane, and the thin red lines denotes when haloes represent 5, 10, and 20 σ over-density fluctuations.

3.4.6 Limitations of our treatment

Though intriguing and possible in principle, the speculations discussed in Section 3.4.4 rely on results strongly dependent on the assumed model for the quasi-star rotation. We therefore comment on the limitations of this model. We model the rotation inside the convective envelope of a quasi-star using the model proposed by Balbus et al. (2009) and Balbus & Weiss (2010). Despite the remarkable agreement with the available data of the internal rotation in the solar convective zone and the physical argumentations supporting its reliability, there are no *a priori* reasons why this model should apply within a quasi-star nor it should produce a sensible description of its rotation, specially at its centre. However, we can test the fundamental assumption behind it, namely that convective cells are long lived compared to the rotation period $t_{\rm rot} \sim 2\pi(\epsilon\Omega_{\rm K,\star})^{-1}$. Since convection produces subsonic motions without net mass redistribution, a rough lower limit for the lifetime of a convective element could be $t_{\rm conv} \sim R_\star/c_{\rm s,0}$. However, we can obtain a better estimate by applying the mixing length theory (e.g. Böhm-Vitense, 1958), which leads to $t_{\rm conv} \sim \sqrt{\alpha\ h_P/(g\delta)}$, where $\alpha \sim 1-2$ is the mixing length parameter[11] (Asida, 2000; Girardi et al., 2000; Palmieri et al., 2002; Ferraro et al., 2006), h_P is the pressure scale-height, $g \sim GM_\star/R_\star^2$ is the gravitational field, and $\delta = \Delta T/T$ is the relative (positive) deviation of the temperature gradient from the adiabatic one in convective regions. The latter is usually tiny, ranging from $\sim 10^{-5}$ to $\sim 10^{-8}$ in deep convective zones (e.g. Böhm-Vitense, 1992; Chabrier, Gallardo & Baraffe, 2007; Prialnik, 2009), and in fact it justifies the description of convective regions through adiabatic relations. Comparing convection and rotation timescales, we obtain:

$$\frac{t_{\rm conv}}{t_{\rm rot}} \sim \frac{\epsilon\sqrt{\alpha}}{2\pi}\left(\frac{h_P}{R_\star\delta}\right)^{1/2} \sim 3.6\ \epsilon_{0.5}\ \delta_{-5}^{-1/2}\ (h_P/R_\star)_{-2}^{1/2}, \tag{3.67}$$

where $\epsilon = 0.5\epsilon_{0.5}$, $\alpha = 2$, $\delta = 10^{-5}\delta_{-5}$, and $h_P/R_\star = 10^{-2}(h_P/R_\star)_{-2}$, as we typically find $h_P/R_\star \gtrsim 0.01$ in the convective envelope of the models of Section 3.3. This order of magnitude calculation suggests that our model should be reasonably applicable to quasi-stars since the convection timescale is at least comparable or even longer than the rotation period, making convective features long-lived enough to couple with and lie along constant Ω contours.

Our model describes a steady-state configuration (thought to be an average in time), whose velocity field is dominated by the azimuthal component, i.e. rotation itself. However, convective regions in differentially rotating zones might lead to features that this approach cannot capture, such as meridional circulation and convective turbulence (e.g. Browning, Brun & Toomre, 2004; Ballot, Brun & Turck-Chièze, 2007; Browning, 2008; Featherstone & Miesch, 2015). Those processes are though to be relevant, especially when coupled with magnetic fields, to understand the long-term maintenance and the mutual powering of the differential rotation and the magnetic dynamo within the Sun. In our case, they might be relevant in the redistribution of angular momentum within the convective envelope, possibly having an effect on the rotation of the central region.

Finally, we recall that quasi-stars are thought to be accreting at high rates ($\gtrsim 1\ {\rm M}_\odot\ {\rm yr}^{-1}$) from the local environment. That means that quasi-stars may not be steady rotating objects, as assumed by our model. Moreover, accretion from outside may proceed either

---

[11] Mixing-lenght theory assumes that convective cells live and mix over a mean free path $l = \alpha h_P$, where α is a free parameter (Böhm-Vitense, 1958).

from a surrounding disc, especially when small scale turbulence is accounted for (Latif et al., 2013), or in a more disordered fashion from filamentary structures carrying angular momentum with various orientations and amplitudes (Choi, Shlosman & Begelman, 2015). In both cases, gravitational torques from non-axisymmetric features might play a relevant role in influencing the redistribution of angular momentum within the quasi-star envelope, though we cannot explicitly account for that in the present work by assuming a steady-state, temporarily-averaged rotation.

3.5 Summary and conclusions

Quasi-stars represent a possible pathway to form a massive black hole seed with mass $\sim 10^{4-5}$ $\mathrm{M}_{\odot}$ through the super-Eddington growth of embryo black holes inside massive quasi-hydrostatic envelopes. Although their existence is far from being observationally demonstrated, they are a key ingredient of many flavours of the direct collapse scenario since they might represent a fundamental stage in the final evolution of supermassive stars.

First, we focus on the impact of outflows onto the final mass with which the black hole seed would emerge at the end of this rapid growth phase. This was already addressed in Dotan, Rossi & Shaviv (2011) but here we consistently solve the full equations for a radiation dominated wind, including both diffusion and advection of energy in the flow. We find that such winds are ultimately powered by advection luminosity within the flow up to the limit where they become nearly adiabatic (i.e. diffusion luminosity is negligible). We confirm that super-Eddington accretion onto newly born black holes within quasi-stars is likely responsible for vigorous mass loss, which in turn may limit the growth of the black holes. If this is the case, forming massive ($> 10^4$ $\mathrm{M}_{\odot}$) seeds via the quasi-star scenario is therefore more difficult than is generally wished for. In contrast, the observable appearance of quasi-stars is expected to be different. They are luminous ($10^{44}-10^{47}$ erg s^{-1}) blue (effective temperature of ~ 8000 K) objects. In colour, they differ from predictions that ignore mass losses, where temperatures can be up to a factor of two lower (Begelman, Rossi & Armitage, 2008). Their characteristics make them promising targets for JWST, while HST may only have detected rare, relative closer ($z < 10$) objects.

Then, we try to push our understanding forward by considering the role of rotation in the evolution of quasi-stars. Specifically, we have addressed the issue of whether the redistribution of angular momentum inside the convective envelope of a quasi-star in steady rotation may favour the formation of a central accretion disc, in order to guarantee the self-consistency of the model. Interestingly, we find that, at given $M_{\bullet}$, most of the massive quasi-stars might not be able to form a central, rotationally-supported accretion region, while the contrary is true for lower mass quasi-stars, typically living within the Evaporation Strip. This bimodal behaviour could lead to different fates, depending on the mass of the original supermassive star at the collapse of the inner core that leads to the formation of the central embryo black hole. At high masses, the black hole might swallow most of the mass that is still infalling from larger radii without providing enough feedback either to stabilise the structure or to halt the collapse. The central black hole would then accrete a large fraction of the envelop mass, possibly reaching $M_\bullet \sim 10^{4-5}$ $\mathrm{M}_\odot$. On the other hand, less massive envelopes might be able to form a central accretion disc and to reach an equilibrium configuration, i.e. a quasi-star. However, outflows then suppress the growth of the central black hole, leading to $M_\bullet \sim 10^{2-3}$ $\mathrm{M}_\odot$. Our results are therefore intriguing, implying possible alternative outcomes for the formation of supermassive black hole seeds by direct collapse. However, this potential needs to be further scrutinised with detailed numerical simulations, as the limitations of our analytical treatment suggest caution.

3.A Appendix. Review of adiabatic winds

For convenience, we briefly review the main features of an adiabatic wind characterised by the equation of state $p = K\rho^{4/3}$, where K is a constant (for further and more general readings, see e.g. Holzer & Axford 1970). Within the formalism introduced in Section 3.2.1, the equations of such an isentropic wind are:

$$\left(1 - \frac{s}{w}\right) w' = -1 + \frac{4s}{1-x}, \tag{3.68}$$

$$w' + 6s' + 1 = 0. \tag{3.69}$$

These equations can be derived by our starting equations of Section 3.2.1 in the optically-thick limit when $\kappa \to +\infty$ and $L(r) \to 0$. From the equation above we immediately see that the critical (or sonic) point (i.e. where $w = s$), when present, coincides with the singular point where w' can diverge. To avoid that, we require that the left-hand side of equation (3.68) is 0 at the critical point x_c, which implies the relation between x_c and the critical velocity w_c:

$$x_\mathrm{c} = 1 - 4w_\mathrm{c}. \tag{3.70}$$

This equation shows that the maximum speed w_c to have a critical point outside the stellar surface (i.e. $x_\mathrm{c} \geq 0$, which corresponds to a subsonic solution at $R_\star$) is $w_\mathrm{c} \leq 1/4$, or $v_\mathrm{c}^2 \leq GM_\star/(2R_\star)$. On the other hand, a critical point approaching infinity (i.e. $x_\mathrm{c} \to 1$) corresponds to a critical velocity $w_\mathrm{c} \to 0$.

Equation (3.69) describes the conservation of energy and entropy. The request of a non-diverging critical point sets the total energy associated to wind:

$$e \equiv w + 6s + x = 3w_\mathrm{c} + 1. \tag{3.71}$$

This condition, when evaluated at $R_\star$, provides a relation between w_c and the quantities $w_\star$ and $s_\star$. The second condition that allows a full determination of $w_\star$ and $s_\star$ given w_c

is the conservation of mass, which reads as equation (3.1), combined with the relation $\rho/\rho_\star = (s/s_\star)^3$, which comes from the equation of state and the general definition of sound speed in equation (3.25). The final and complete relations between w_c and $w_\star$ and $s_\star$ are:

$$w_\star + 6s_\star = 3w_\mathrm{c} + 1, \tag{3.72}$$

$$16\ w_\star^{1/2} s_\star^3 = w_\mathrm{c}^{3/2}. \tag{3.73}$$

When $w_\mathrm{c} \to 0$, the second relation shows that $w_\star^{1/2} s_\star^3 \to 0$ and either $w_\star$ or $s_\star$ has to go to zero. Since this limit case corresponds to have a critical point well outside $R_\star$, we have $w_\star \ll s_\star$ and therefore $w_\star \to 0$. The first relation then imply that every adiabatic solution with a critical point at a finite radius has necessarily $s_\star > 1/6$, while every solution with a critical point at a finite radius larger than $R_\star$ requires $1/6 < s_\star < 1/4$. Combining equations (3.72) and (3.73) provides also a unique relation $\mathcal{M}_{\star,\mathrm{ad}}(s_\star)$ between the sound speed and the Mach number at $R_\star$. This relation is shown in Figure 3.12.

We can finally estimate the mass outflow rate through the conditions at the critical point and map them back to the properties of the flow at $R_\star$. At the critical point, the outflow rate reads $\dot{M} = 4\pi r_\mathrm{c}^2 \rho(r_\mathrm{c}) v_\mathrm{c}$. Using the relations $r_\mathrm{c} = GM_\star/(2v_\mathrm{c}^2)$, $\rho(r_\mathrm{c}) = \rho_\star (v_\mathrm{c}/c_{\mathrm{s},\star})^6$ and equation (3.72) in dimensional form, we obtain:

$$\dot{M}_\mathrm{ad} = \frac{4\pi G^2 M_\star^2 \rho_\star}{\sqrt{2} c_{\mathrm{s},\star}^3} \left(1 + \frac{\mathcal{M}_{\star,\mathrm{ad}}^2(c_{\mathrm{s},\star})}{6} - \frac{v_\mathrm{esc}^2}{6 c_{\mathrm{s},\star}^2} \right)^{3/2}. \tag{3.74}$$

This depends on the properties of the star, such as $M_\star$ and v_esc, on the density $\rho_\star$ that sets the normalisation, and on the sound speed $c_{\mathrm{s},\star}$. Figure 3.12 shows also the scaling of $\dot{M}_\mathrm{ad}$ with the sound speed $c_{\mathrm{s},\star}$.

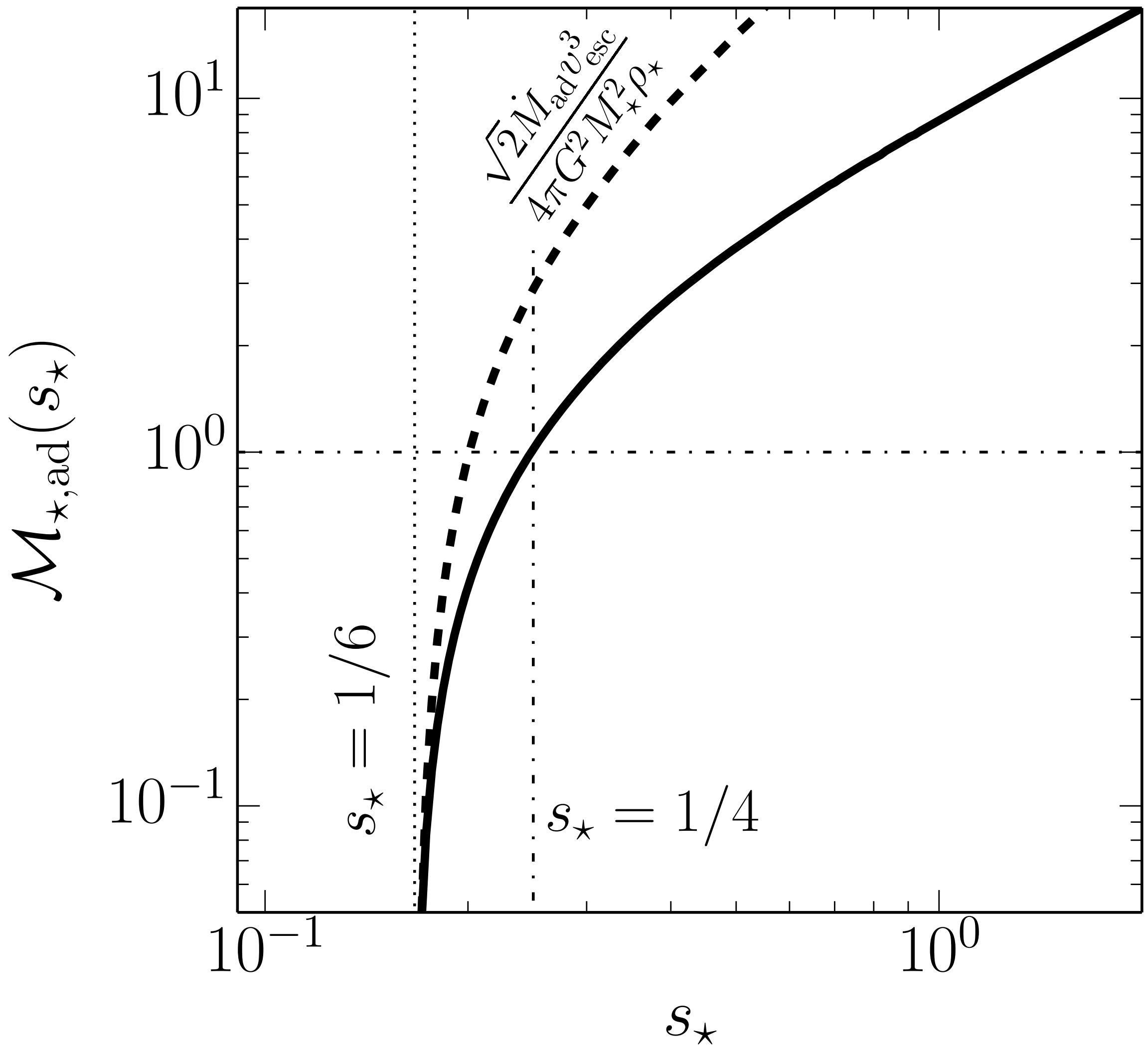

Figure 3.12 Relation $\mathcal{M}_{\star,\mathrm{ad}}(s_\star)$ between the sound speed and the Mach number at $R_\star$ for an adiabatic wind (thick continuous line). The thick, dashed line shows the scaling of the mass outflow rate $\dot{M}_{\mathrm{ad}}$ with the sound speed $c_{\mathrm{s},\star}$. The thin, dotted line shows the position $s_\star = 1/6$, where the Mach number approaches asymptotically zero. The thin, dot-dashed lines show that when $s_\star = 1/4$, the wind has $\mathcal{M}_\star = 1$ exactly at $R_\star$.

4

The evolution of massive black hole pairs in clumpy environments[1]

Abstract In this Chapter I will discuss the evolution of massive black hole pairs within gaseous environments characterised by highly inhomogeneous and clumpy distributions of mass (e.g. massive clouds or spiral arms). By means of hydrodynamical simulations, I will show that several dynamical processes can affect the dynamics of the pairs: (i) the black holes can undergo two-body-like interactions with massive clumps; (ii) they can be scattered through gravitational torques induced by spiral arms; and (iii) they can tidally disrupt and accrete mass from clouds, increasing their effective mass. The consequence of that, which has been tested with idealised simulations of gaseous circumnuclear discs as well as self-consistent galaxy mergers, is that the orbital decay timescale from separations ~ 500 pc can stochastically either increase or decrease by up to an order of magnitude, compared to the unperturbed case. I discuss the implications of this finding.

---

[1] The material of this chapter is contained in and adapted from the following publications: **Fiacconi**, Mayer, Roškar, & Colpi 2013, ApJ, 777, L14; Roškar, **Fiacconi**, Mayer, et al. 2015, MNRAS, 449, 494.

4.1 Introduction

After the discovery of the scaling relations between the mass of supermassive black holes and some of the properties of the host galaxies (e.g. Magorrian et al. 1998; Ferrarese & Merritt 2000; Tremaine et al. 2002; Marconi & Hunt 2003; Gültekin et al. 2009; McConnell & Ma 2013; see also Section 1.2.2 for further details), supermassive black holes have been shown to be ubiquitous in hosts down to disc-galaxy mass scales in the local Universe (Ferrarese & Ford, 2005; Greene et al., 2010), as well as at high redshift (Schawinski et al., 2012). There is also growing evidence supporting the presence of supermassive black holes in dwarf galaxies with stellar masses $\lesssim 10^9$ $M_\odot$ (Reines et al., 2011; Reines, Greene & Geha, 2013; Koss et al., 2014).

Combining this piece of evidence with the Λ-CDM hierarchical merging paradigm of structure formation (see Section 1.1), it implies that supermassive black hole pairs should be present in a wide range of merging galaxies (see Section 1.2.5). When two galaxies merge, the dynamics is initially dominated by dynamical friction against the dark matter background that makes their separation shrink from > 10 kpc down to ~ 1 kpc. During this phase, the black holes are bound to their galactic cores and they may experience accretion episodes triggered by gravitational perturbations and rapid time variations of the local potential wells. In fact, they are often observed as offset or dual AGN when one or both black holes shine, respectively (Komossa et al., 2003; Liu et al., 2011; Comerford et al., 2013; Liu et al., 2013; Comerford & Greene, 2014; Comerford et al., 2015). Afterward, the evolution of the pair is still dominated by dynamical friction, but of the supermassive black holes (possibly with the surrounding mass that is bound to them) against the stellar and/or gaseous background, until they form a bound binary at ~ 1 pc separation. Then, they may eventually coalesce through the emission of gravitational waves, when the interactions with either gas or stars can shrink their separation further.

However, the latter phase is still not completely understood, and in particular the timescale of the orbital decay is still unclear. It has been shown that a binary could stall at ~ 1 pc separation because of the depletion of the loss-cone[2], leading to the so called "last parsec problem" (Milosavljević & Merritt, 2001, 2003; Makino & Funato, 2004). However, it has been also suggested that the black holes can actually merge within an Hubble time when the triaxiality and/or the rotation of the galactic host is realistically taken into account (Berczik et al., 2006; Sesana, Haardt & Madau, 2007; Khan, Just & Merritt, 2011; Khan et al., 2013; Holley-Bockelmann & Khan, 2015; Vasiliev, Antonini & Merritt, 2015).

The evolution of a massive black hole pair changes when the merger involves gas-rich galaxies. Simulations of major mergers between gas-rich disc galaxies show that gas can be

---

[2] The loss cone is the region in phase-space that characterises low angular momentum orbits able to plunge all the way to the centre, where the binary likely resides.

effectively funnelled at the centre of the remnant where it likely forms a $\gtrsim 10^8$ $M_\odot$ circumnuclear disc ~ 100 pc in size (Di Matteo et al., 2007; Mayer et al., 2007; Mayer, Kazantzidis & Escala, 2008; Sijacki, Springel & Haehnelt, 2011; Chapon, Mayer & Teyssier, 2013), as confirmed by observations (Sanders & Mirabel, 1996; Downes & Solomon, 1998; Davies, Tacconi & Genzel, 2004; Medling et al., 2014). As a consequence, the dynamics of the black hole pair when they are embedded in the galactic disc is dominated by the dynamical friction against the dense gas and by gravitational torques from non-axisymmetric features (Escala et al., 2004, 2005; Mayer, 2013). In such conditions, the orbital decay of the black hole pair down to ~ 1 pc can be very rapid ($\ll 1$ Gyr), with mild sensitivity to the orbital and geometric parameters of the merger (Kazantzidis et al., 2005).

As the merger leads to a central inflow of gas, such gas can experience episodes of intense star formation (i.e. merger-driven nuclear starburst; Downes & Solomon 1998; Smith et al. 2007; Smirnova & Moiseev 2010; Beletsky et al. 2011). Under these conditions, the interstellar medium of the circumnuclear disc and the surrounding is expected to be clumpy and inhomogeneous on scales of a few to several tens of parsecs, owing to gravitoturbulent fragmentation governed by the Toomre instability (Agertz et al., 2009; Tasker & Tan, 2009; Klessen et al., 2010). Most of the previous investigations have assumed that the self-gravitating gas in the circumnuclear disc is warm enough to avoid fragmentation; however, a realistic description of a nuclear starburst requires to take into account the inhomogeneous and clumpy structure of the multi-phase interstellar medium. Therefore, we proceed to address the question as to the black hole pair dynamics is affected by such environmental conditions.

Content of the Chapter In the following, we model two progressively less idealised configurations. First, we start in Section 4.2 by assuming that a binary merger between gas-rich galaxies has occurred, and the two black holes have been able to sink within the inner kiloparsec. Building upon the results of previous calculations (Escala et al., 2005; Mayer et al., 2007; Chapon, Mayer & Teyssier, 2013) and recent observations (Medling et al., 2014), the black holes will likely be embedded in a rotationally-supported circumnuclear gaseous disc. This disc is a few hundreds of parsecs in size and $\sim 10^8$ $M_\odot$ in mass, and it sits at the centre of the bulge of the merger remnant. The disc is forces to develop an inhomogeneous and multiphase interstellar medium by adopting a simple cooling prescription to maintain a gravitoturbulent state. Then, we continue in Section 4.3 by self-consistently forming the circumnuclear disc within a full binary merger between massive spiral galaxies at low redshift, following the dynamics of the black hole pair at gradually higher resolution down to parsec separations. We finally discuss and summarise the results in Section 4.4.

4.2 Evolution of massive black hole pairs in circumnuclear discs

4.2.1 Set-up of the numerical simulations

We perform a suite of numerical simulations to study the evolution of a massive black hole pair embedded in a gaseous circumnuclear disc which, in turn, is at the centre of a stellar spheroid. All the simulations were run with the Tree/smoothed particle hydrodynamics (SPH) N-body code GADGET2 (Springel, 2005). Our models resemble the inner region of the remnant of a merger that involved two galaxies with a central massive black hole each, in a fashion similar to Escala et al. (2005) and Dotti et al. (2007, 2009). The gaseous disc has a Mestel surface density profile with a scale length $R_{\mathrm{d}} = 100$ pc and a radial extent of ~ 150 pc. The vertical structure is initially Gaussian, with a scale height $z_{\mathrm{d}}(R) = h\ R$ and aspect ratio $h = 0.05$. The gas has an initial uniform temperature $T_0 = 20,000$ K. The disc is embedded in a Plummer stellar spheroid that represents the innermost part of a bulge, with scale radius $r_\star = 50$ pc and radial extent of ~ 500 pc. Among the different runs, we vary the mass of the spheroid $M_\star$ and the mass of the disc M_{d}, but we fix the ratio $M_\star/M_{\mathrm{d}} = 5$, in fair agreement with observations (Downes & Solomon, 1998). We place the first black hole of mass $M_{\bullet 1} = 10^7\ \mathrm{M}_\odot$ at the centre of the disc and we let the models relax for 10 Myr toward equilibrium. We assume a polytropic equation of state for the gas with adiabatic index $\gamma = 1.4$, in agreement with previous theoretical (Klessen, Spaans & Jappsen 2007; Mayer et al. 2007, 2010) and observational (Downes & Solomon, 1998) work that has studied the conditions of the central gas in ongoing mergers or merger remnants.

Then, we finally added the second black hole of mass $M_{\bullet 2} = q\, M_{\bullet 1}$ at the initial separation $a_0 = 60$ pc from the central one. Both black holes are treated as collisionless particles. The orbit of the secondary black hole is specified by the ratio f between the radial and the azimuthal components of the initial velocity $\mathbf{v}_0$, with the constraint $|\mathbf{v}_0| = V_{\mathrm{c}}(a_0)$, where $V_{\mathrm{c}}(a_0)$ is the circular velocity in a_0. f also specifies the initial eccentricity e_0 of the orbit, $e_0 \sim \sqrt{1 - 1/(1 + f^2)}$. All the models are composed of 2×10^5 SPH particles and 1×10^6 collisionless particles for the gaseous disc and the stellar spheroid, respectively. This corresponds to a mass resolution m_{p} that varies between 500 and 2500 $\mathrm{M}_\odot$, depending on the mass of the disc and the spheroid. The force resolution set by the gravitational softening is $\epsilon_{\mathrm{g}} = 0.5$ pc for all the particles.

For each choice of parameters M_{d}, q and f, we initialise three sets of initial conditions. In the first one, which we will refer to as the "smooth" set, the secondary black hole is added immediately after the 10 Myr relaxation phase. In the other two, we added the secondary black hole after we forcefully make the disc Toomre unstable by cooling it during an additional relaxation phase via the phenomenological cooling term:

$$\Lambda_{\rm cool} = -\frac{u}{t_{\rm cool}}, \tag{4.1}$$

where u is the specific internal energy of the gas and $t_{\rm cool}$ is a constant cooling timescale. We chose different parameters to create two different environments: in the first case ("clumpy 1"), we further relaxed the disc for 2 Myr with $t_{\rm cool} = 0.2$ Myr. The disc becomes violently unstable and fragments in many small and dense clumps that account for $\gtrsim M_{\rm d}/2$. In the second case ("clumpy 2") we used $t_{\rm cool} = 1$ Myr for 10 additional Myr and we turn off the cooling locally for densities $> 5 \times 10^5$ H cm^{-3}. The disc has a more developed background component compared to "clumpy 1" models, with fewer, less concentrated clumps.

Label	$M_{\rm d}$ (M$_\odot$)	q	f	e_0	$t_{\rm cool}$ (Myr)
q005f02LM	10^8	0.05	0.2	0.2	1.0
q005f1LM	10^8	0.05	1.0	0.7	1.0
q02f025LM	10^8	0.2	0.25	0.25	1.0
q02f2LM	10^8	0.2	2.0	0.9	1.0
q01f02HM	5×10^8	0.1	0.2	0.2	0.5
q01f2HM	5×10^8	0.1	2.0	0.9	0.5
q02f02HM	5×10^8	0.2	0.2	0.2	0.5
q02f2HM	5×10^8	0.2	2.0	0.9	0.5

Table 4.1 List of performed simulations and of their parameters. From left to right: label of the simulation, mass of the disc, mass ratio of the black holes $q = M_{\bullet 2}/M_{\bullet 1}$, where $M_{\bullet 1} = 10^7$ M$_\odot$, ratio f between the radial and the azimuthal initial velocity components of $M_{\bullet 2}$ (see the text for details), initial eccentricity $e_0 \sim \sqrt{1 - 1/(1+f^2)}$ of the secondary black hole, cooling time.

Table 4.1 lists all the simulations and summarises the adopted parameters. The values of $t_{\rm cool}$ for both "clumpy 1" and "clumpy 2" models are chosen to be comparable to the dynamical time of the discs at $\sim R_{\rm d}$ to allow gravitoturbulent fragmentation to set in and sustain. We keep the density switch for the cooling during "clumpy 2" simulations even after we inserted the secondary black hole. We stress that equation (4.1) represents a phenomenological way to create and maintain strong inhomogeneities in the CND.

4.2.2 Results

Orbital decay in a smooth disc: overview

The secondary black hole of all the "smooth" simulations moves toward the disc centre on a typical timescale of $\sim$ 10 Myr. This is shown in Figure 4.1, which compares the time evolution of the separation between the two black holes for all the corresponding "smooth", "clumpy 1" and "clumpy 2" simulations. The orbital decay of the secondary

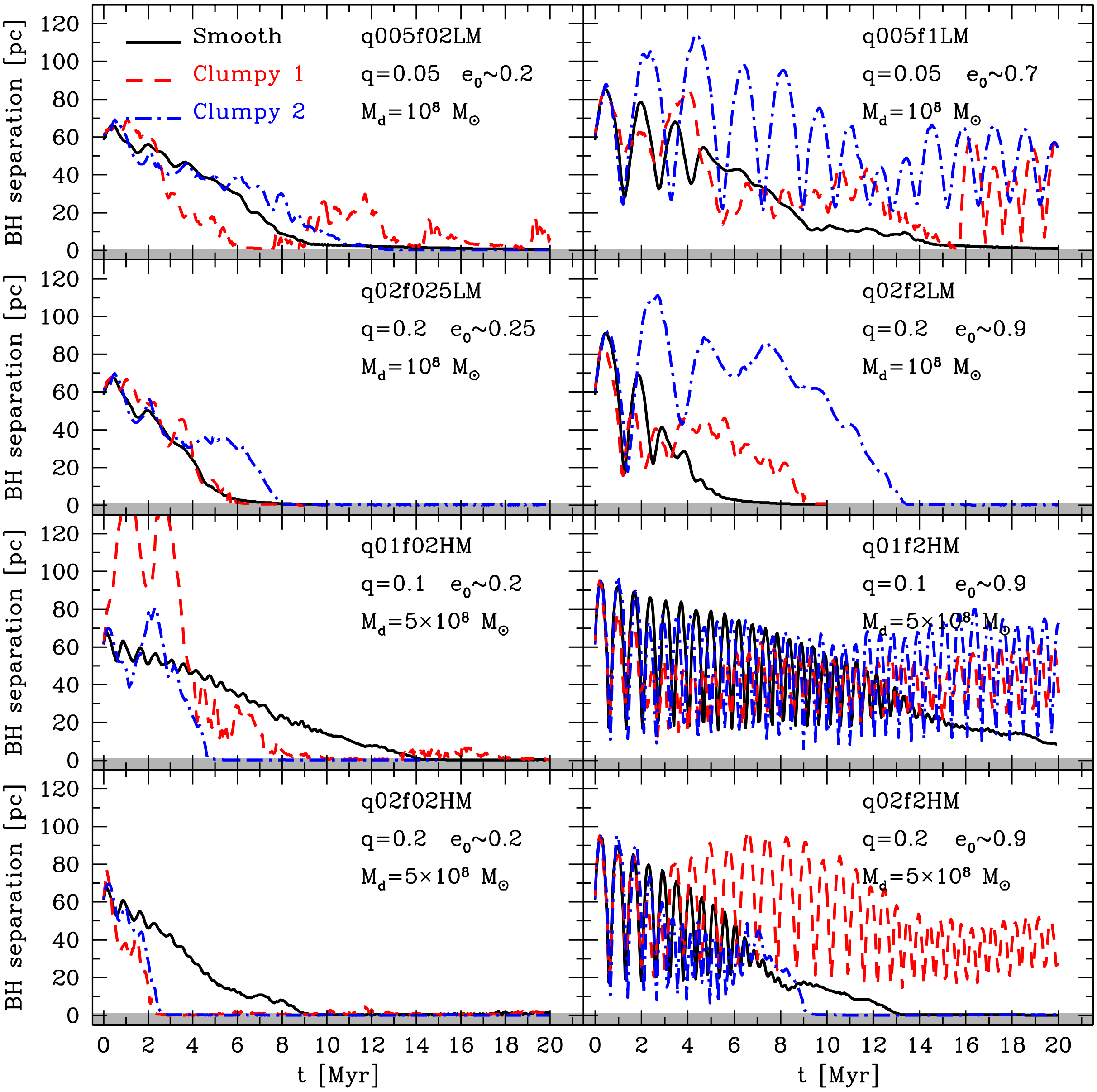

Figure 4.1 Time evolution of the black hole separation for all the corresponding "smooth", "clumpy 1" and "clumpy 2" models. The top two rows show runs with $M_\mathrm{d} = 10^8$ M$_\odot$, whereas the bottom two with $M_\mathrm{d} = 5 \times 10^8$ M$_\odot$. For each pair of rows, the first and the second one shows runs with lower and higher q, respectively. The left and the right column shows runs with lower and higher e_0, respectively. Black continuous lines, red dashed lines and blue dot-dashed lines show the black hole separation of the "smooth", "clumpy 1" and "clumpy 2" models, respectively.

black hole in the "smooth" discs is characterised by two phases. Initially, the black hole induces a trailing hydrodynamical wake that, in turn, makes the orbit circularise (Dotti et al., 2007) because of conventional dynamical friction (Chandrasekhar, 1943; Ostriker, 1999; Colpi, Mayer & Governato, 1999). This is shown in Figure 4.2 for the first two orbits of the q005f1LM run. When the orbit is close to circular, the relative velocity between the black hole and the gas becomes low. This causes a change in the orbital decay timescale. The black hole induces a density wave perturbation, eventually amplified by the self-

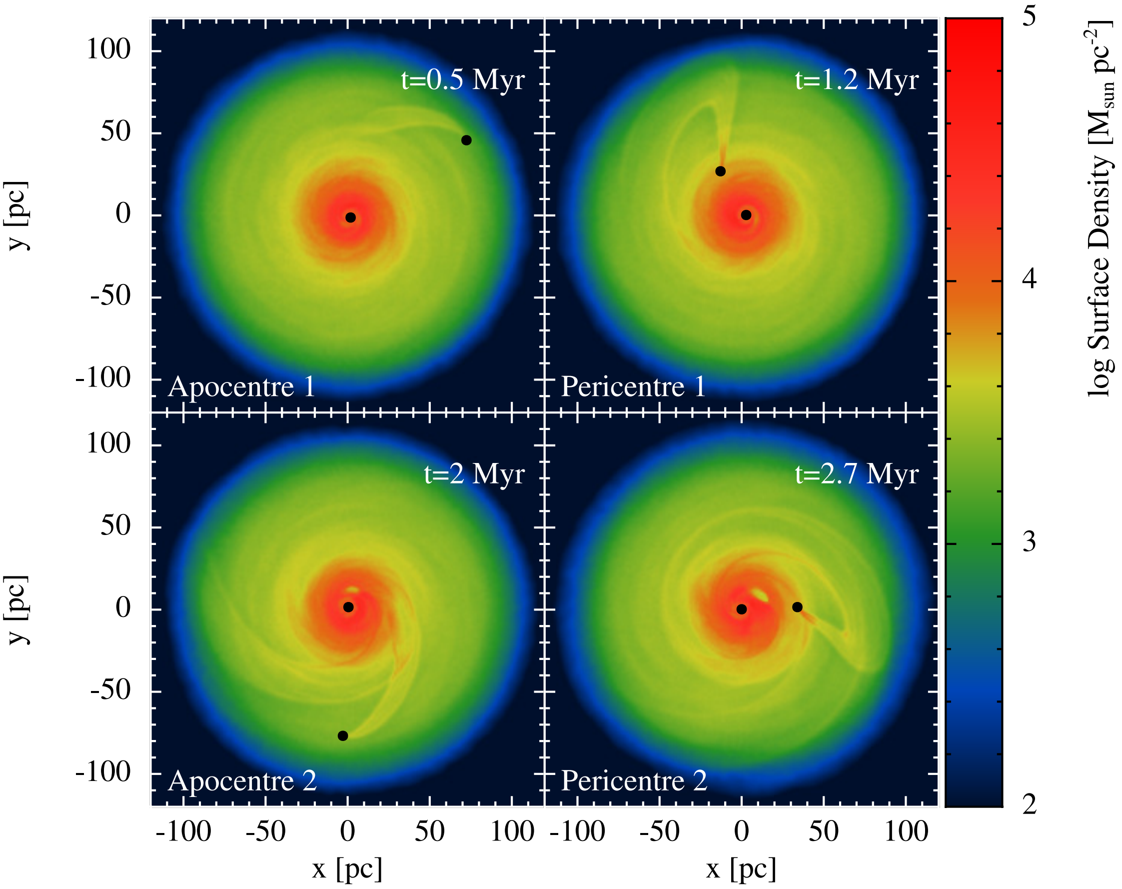

Figure 4.2 Time evolution of the gas surface density of the "smooth" model of run q005f1LM. The size of each panel is 240 pc per edge. The position of the black holes are marked by two white dots. The panels show the first two apocentres once pericentres, when the dynamical friction wake is in front of and behind the black hole, respectively.

gravity of the disc that exerts a global torque on the perturber itself, initiating its rapid sinking toward the centre. The latter phase resembles the regime of standard Type I planet migration (Lin & Papaloizou, 1986). An aspect that has not been appreciated in the literature is that the secondary black hole's angular momentum loss occurs on an intrinsically shorter timescale during the latter phase, as shown in Figure 4.3.

Furthermore, while dynamical friction shuts off at separations such that the enclosed mass within the pair's orbit is $\sim M_{\bullet 1}+M_{\bullet 2}$, global torques will continue to act as long as a sufficiently massive gaseous disc is present at large radii and the black hole does not open a gap (Chapon, Mayer & Teyssier, 2013). Finally, in addition to the disc contribution to the orbital decay, the secondary black hole also suffers dynamical friction exerted by the stellar spheroid. However, the stellar torque is weaker than the torque that comes from the gaseous disc, which mainly drives the dynamical evolution of the pair.

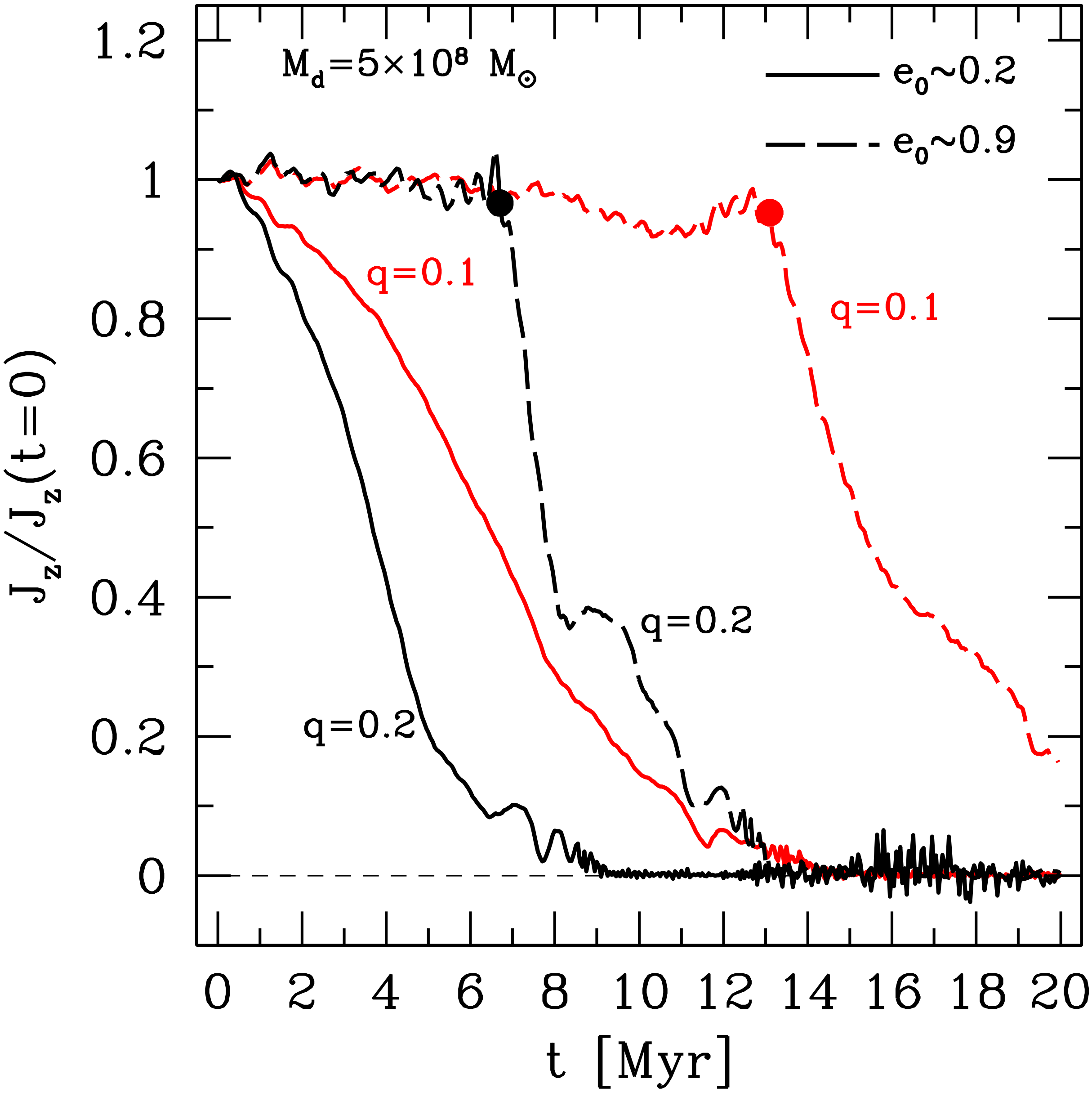

Figure 4.3 Time evolution of the secondary black hole angular momentum for the "smooth" models with $M_\mathrm{d} = 5 \times 10^8$ $\mathrm{M}_\odot$. Solid and dashed lines represent runs with $e_0 \simeq 0.2$ and $e_0 \simeq 0.9$, respectively. Black and red colours represent runs with $q = 0.1$ and $q = 0.2$, respectively. The dots qualitatively mark the transition between the two phases described in the text.

Black hole pair evolution in a clumpy disc

The orbital evolution of the secondary black hole in a clumpy disc is affected by the dynamical interaction with clumps acting as massive perturbers. Figure 4.4 shows the gas surface density of the disc for the "clumpy 1" model of run q02f2HM. The disc is sprinkled with massive clumps and non-axisymmetric features that exert torques on the secondary black hole. Figure 4.5 shows the normalised initial mass function (just after the relaxation phase) and the mass-size relation of the clumps of "clumpy 1" and "clumpy 2" models. We compare our data with the observed sample of molecular clouds in the Galactic centre taken from Oka et al. (2001). Although the Galactic centre likely represents a less extreme environment than the inner region of a merger remnant that we aim at modelling, our models are in fair agreement with the observational data and are only weakly influenced by the value of $t_{\rm cool}$ adopted in the relaxation phase.

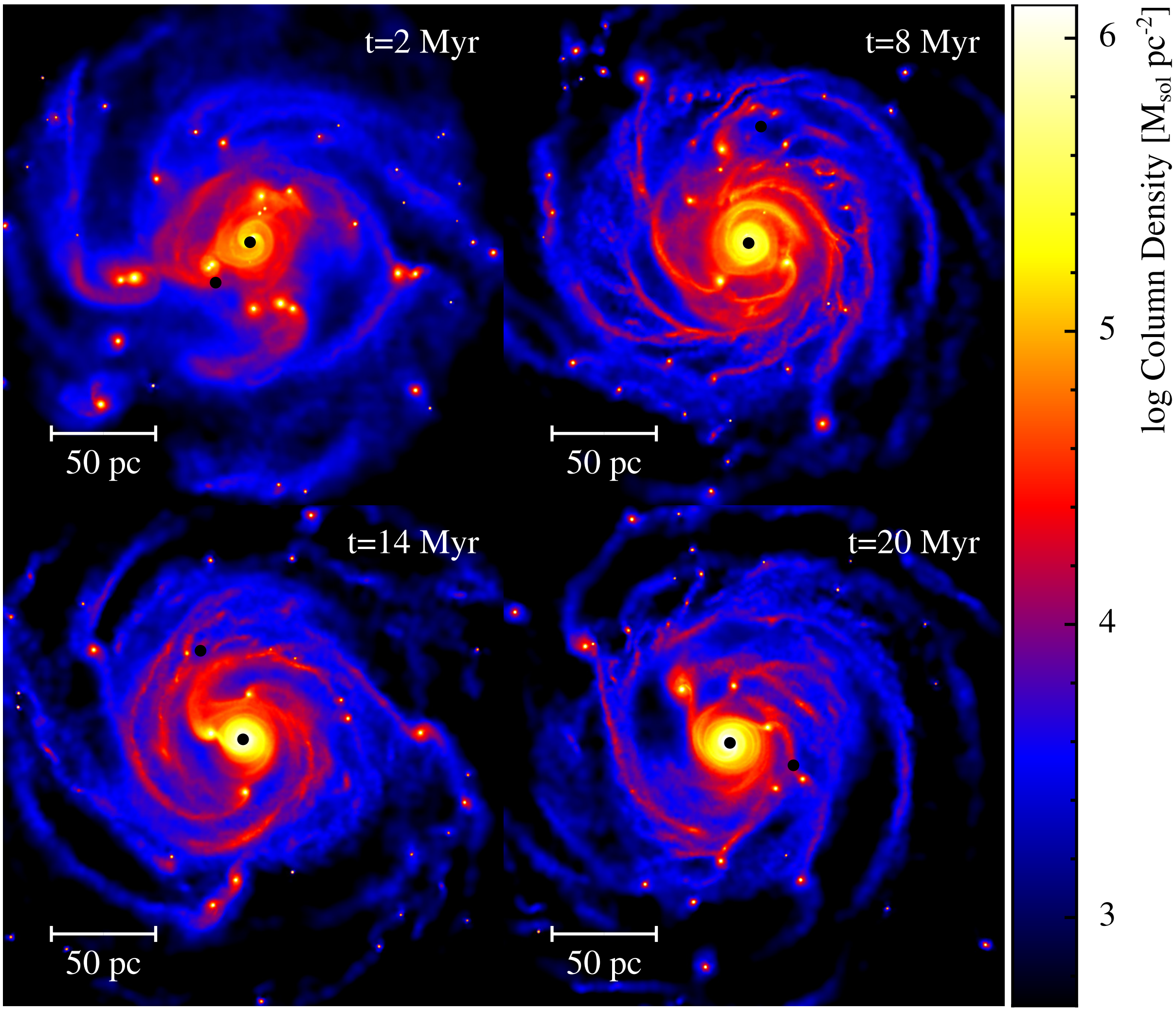

Figure 4.4 Time evolution of the gas surface density of the "clumpy 1" model of run q02f2HM. The size of each panel is 240 pc per edge. The position of the black holes are marked by two black dots.

Close encounters (at separation of $\sim 3-10$ pc) between the secondary black hole and the massive clumps occur in both "clumpy 1" and " clumpy 2" models. black hole-clump interactions act as gravitational slingshots, causing an exchange of orbital energy and angular momentum. These impulsive perturbations make the black hole deviate from its original orbit, shifting its mean orbital radius either inward or outward.

Gravitational slingshots can delay the sinking of the secondary black hole when they cause its temporary ejection from the disc plane. This happens both with the dense clumps of "clumpy 1" models and with the more diffuse ones of "clumpy 2" models, but

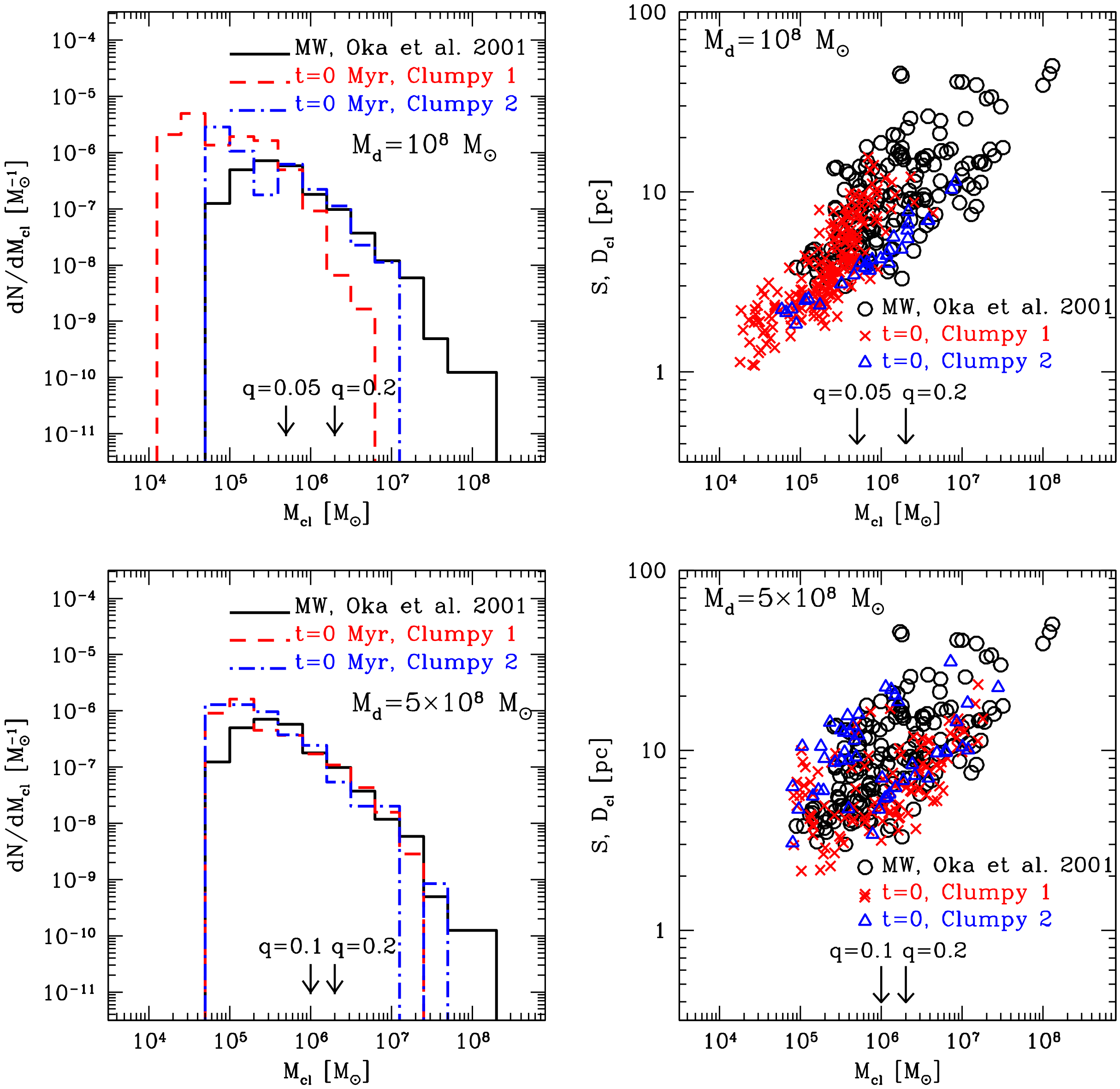

Figure 4.5 Top-left panel: normalised mass function of clumps for “clumpy 1” and “clumpy 2” models with $M_\mathrm{d} = 10^8$ M$_\odot$. Bottom-left panel: same as before for models with $M_\mathrm{d} = 5 \times 10^8$ M$_\odot$. Top-right panel: clump mass vs. size (S from Oka et al. (2001) and clump diameter D_cl) for models with $M_\mathrm{d} = 10^8$ M$_\odot$. Bottom-right panel: same as before for models with $M_\mathrm{d} = 5 \times 10^8$ M$_\odot$. Black continuous lines (circles), red dashed lines (crosses) and blue dot-dashed lines (triangles) show the data from Oka et al. (2001), “clumpy 1” and “clumpy 2” models, respectively. Vertical arrows indicate the masses of the secondary black holes.

preferentially for high-e_0 runs (e.g. run q005f1LM, q02f2LM and q01f2HM). After the interaction, the secondary black hole moves on an eccentric orbit, tilted up to $\sim 15° - 40°$ with respect to the original plane for both “clumpy 1” and “clumpy 2” models. The black hole then spends most of the orbital period moving outside the disc, where is mostly subjected to dynamical friction against the stellar background and the decay timescale becomes longer, in agreement with Escala et al. (2005). The actual value of the decay timescale can depend substantially on the density profile of the stellar spheroid. Sinking black holes can stall in constant density cores when they approach the core radius, as shown by Read et al. (2006).

Clumps have also an indirect retarding effect on the black hole orbital decay, especially for eccentric orbits. They produce spiral arms that exert torques on the secondary black hole and perturb the background component, stifling the coherence of the black hole hydrodynamical wake that would lead to the circularisation of the orbit and the subsequent efficient decay phase.

On the other hand, either the black hole can be captured inside a massive clump, or a clump can be tidally heated and then disrupted and accreted by the orbiting black hole during a close passage at low relative velocity. In both cases, the black hole ends up embedded in a gaseous cloud bound to it. Although this cloud cannot be actually accreted due to the lack of an accretion prescription in our simulations, it provides an higher effective mass that reduces the orbital decay timescale[3] up to ~ 2 Myr (e.g., "clumpy 2" model of run q01f02HM and q02f02HM).

4.3 Evolution of massive black hole pairs in galaxy mergers

4.3.1 Set-up of the numerical simulations

Simulation Code

Our simulations have been run using the SPH code GASOLINE (Wadsley, Stadel & Quinn, 2004), which is an extension of the N-body/Tree code PKDGRAV (Stadel, 2001). All analysis was performed using the open source PYNBODY package (Pontzen et al., 2013). The simulation uses standard prescriptions for star formation from Stinson et al. (2006) using a Salpeter initial mass function (Salpeter, 1955). Star particles are spawned from a gas particle whose density $\rho > \rho_{\rm thresh}$ and $T < T_{\rm thresh}$ at the rate $\dot{\rho}_\star = \epsilon_{\rm SF}\rho/t_{\rm dyn} \propto \rho^{3/2}$, where $t_{\rm dyn} = 1/\sqrt{G\rho}$ is the local dynamical time and $\epsilon_{\rm SF} = 0.01$ is the star formation efficiency. $\rho_{\rm thresh}$ and $T_{\rm thresh}$ are density and temperature thresholds chosen such that the stars are forming in the densest, coldest gas regions found in the simulation. Due to the large dynamic range spanned by the simulation from the initial galaxy-scale stages to the final stages where we focus only on the inner regions, we have had to adjust these star formation parameters at different times. Initially, when the two merging systems are still distinct, we use a star formation density threshold $\rho_{\rm thresh} = 0.1$ H cm^{-3} and a temperature threshold $T_{\rm thresh} = 1.5 \times 10^4$ K. However, as the simulation progresses and the gas phase includes low-temperature, high-density material, we revise this prescription to $\rho_{\rm thresh} = 10^4$ H cm^{-3} and $T_{\rm thresh} = 200$ K. The star-formation parameters are adapted

---

[3] We recall that both Type I migration and dynamical friction torques that provide loss of angular momentum and/or orbital energy scale with the perturber mass squared.

to the new resolution when particle splitting is performed. Due to the short orbital times inside the nuclear region we reduce the star formation timestep $\Delta t_{\rm SF}$ (the minimum time between two episodes of star formation) to 10^5 yr, while prior to the splitting this is $\sim 10^6$ yr, as typically done in galaxy scale simulations. This enforces an almost continuous age distribution for stellar particles across time, even in the most rapidly evolving regions of the simulation.

The supernova feedback is the "blastwave" feedback from Stinson et al. (2006), which attempts to model the expansion of supernova-driven bubbles by mimicking the ballistic phase of the shock triggered by the supernova explosion. To mimic this phase, cooling is turned off for the timescale of the snowplow phase of the shock, calculated based on the instantaneous SN energy input and the ambient density following the analytical model of McKee & Ostriker (1977). Such a feedback prescription overcomes the difficulty in distributing SN energy radiatively in the interstellar medium; due to the implicit optically thin gas modelling and the high densities involved, the cooling timescales would be otherwise extremely short. The blastwave feedback allows us to efficiently couple the SN energy to the interstellar medium. In addition, the feedback model pollutes the gas with metals produced in SN Ia, SN II, and AGB star winds (see Stinson et al. 2006 for more details).

In the simulations presented in this paper we only consider cooling by hydrogen and helium (GASOLINE does compute non-equilibrium rates at any given step rather than simply assuming a cooling function), which implies the temperature of the gas levels off at $\sim 10^4$ K (see e.g. Sutherland & Dopita 1993). We ignore contributions to the cooling function from metals above 10^4 K, which could have an impact on returning the gas expelled via feedback back to the central region. However, even with metal-line cooling, we expect the cooling timescales for 10^6 K coronal gas to be much longer than the supermassive black hole coalescence timescale (of order a few million years based on Mayer et al. 2007 and Chapon, Mayer & Teyssier 2013). Therefore, while not including the metal-dependent cooling may certainly influence the long-term post-merger evolution, we don't expect for it to significantly alter the properties of the rather short nuclear disc rebuilding phase covered in this paper. We allow the gas to cool below 10^4 K including metal lines using the empirical fit from Mashchenko, Wadsley & Couchman (2008) based on calculations by Bromm et al. (2001) until the gas is optically-thin to stellar radiation, and then switch to a novel thermal balance prescription for high density optically thick gas, as described next.

Thermodynamics of the high density gas phase

The gas equation of state plays a decisive role in the fate of supermassive black hole binaries (Mayer et al., 2007; Mayer, Kazantzidis & Escala, 2008; Chapon, Mayer & Teyssier, 2013). In particular, the behaviour of the cold, high density, optically thick gas phase was poorly modelled in previous simulations adopting a prescribed equation of state. Here, we improve on previous work by including star formation in the coldest, densest gas, as well as stellar feedback, during all stages of the simulation. In addition, we implement a table for equilibrium temperatures based on Spaans & Silk (2000) appropriate for $\rho_{\rm gas} > 0.1$ H cm^{-3}. The table gives an equilibrium temperature given a gas density following a detailed calculation including the relevant radiative processes in the densest gas phase. The model has been calibrated using 2D radiative transfer calculations for irradiated clouds in starburst environments (Spaans & Silk, 2000). These include: (i) stellar UV heating on dust and IR dust emission, (ii) photoelectric heating effect on dust, (iii) cosmic ray heating, (iv) trapping of molecular and atomic lines in presence of a photodissociation layer, and (v) local turbulent velocity dispersion. The model is essentially an upgraded version of that adopted in Klessen, Spaans & Jappsen (2007). It assumes a star formation rate of 100 $\rm M_{\odot}$ yr^{-1} when computing the UV flux from stars and ionisation equilibrium between species.

In practice, we implement this cooling table as a temperature correction on top of the usual temperature calculation. If the particle's density is in the range described by the table, then we modify its temperature $T_{\rm p}$ by a factor $\Delta T = (T_{\rm eq} - T_{\rm p})/T_{\rm eq}$, where $T_{\rm eq}$ is the equilibrium temperature interpolated from the cooling table. In this way, the particles are pushed toward the equilibrium solution rather than simply assigned a temperature. Figure 4.6 shows the temperature evolution of a single particle at different densities, for solar (top panel) and super-solar metallicity (bottom panel). The dashed lines show the cooling trajectories using our temperature correction, while the solid lines show the standard low temperature cooling. One can see clearly that the temperature correction initially accelerates the cooling, but as the gas gets colder it introduces a higher temperature floor at a given density. Essentially, this allows us to capture the reduced efficiency of cooling in regions of high optical depth in the very dense regions of our system, without having to resort to a full radiative transfer calculation. Note that the temperature correction is applied after the energy calculation, so as to make sure that the cooling still takes place at a reasonable rate. In Figure 4.7 we show the phase diagram using this low temperature cooling correction. We chose an output at the moment of the final apocentre, when the softening is reduced to 1 pc to ensure that the high-density end of the distribution is well-populated. The orange and dashed blue lines show the modified cooling table values for $Z_{\odot}$ and $3Z_{\odot}$ respectively.

Since we are concerned with modelling the inhomogeneous interstellar medium and we are allowing the gas to cool to low temperatures, we must take care to guarantee that any fragmentation remains physical. We ensure that the gas particle's implied Jeans scale is resolved at a given temperature and density by imposing a pressure floor constraint following Agertz, Teyssier & Moore (2009). The minimum pressure for each individual

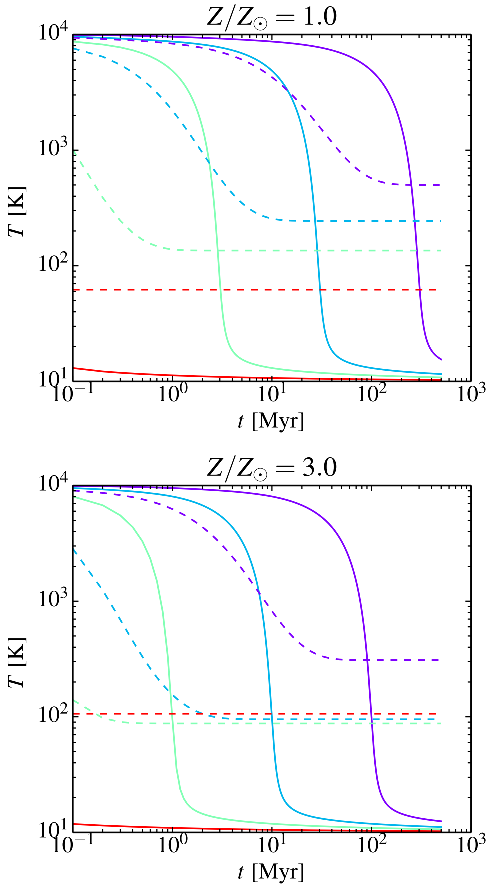

Figure 4.6 Temperature evolution of a single particle using the standard low-temperature cooling curve (solid lines) and our temperature correction using calculations from Spaans & Silk (2000) (dashed lines). Purple, turquoise, green and red colours represent densities of 1, 10, 100 and 10^4 H cm^{-3} respectively. Upper panel shows the case for metallicity $Z = Z_\odot$, while the lower panel show the case for $Z = 3Z_\odot$.

gas particle is set to $P_{\rm min} = \alpha {\rm max}(\epsilon, h)^2 G\rho^2$, where $\alpha = 3.0$ is a safety factor, G is the gravitational constant, ϵ is the softening length, h is the smoothing length and ρ is the particle density. In Figure 4.8 we show the distribution of the ratios of $M_{\rm jeans}/M_{\rm part}$ during the final part of the simulation when the force resolution is 1 pc, showing that the Jeans mass is resolved by $\gtrsim 10$ particles everywhere in the simulation and therefore any fragmentation and clumping we see is physical.

Particle Resampling during the Galaxy Merger

The system is initialised as a merger between two Milky Way-like discs each with a supermassive black hole particle embedded in the centre. The same initial conditions were used as a starting point for the simulations presented in Mayer et al. (2007). Here

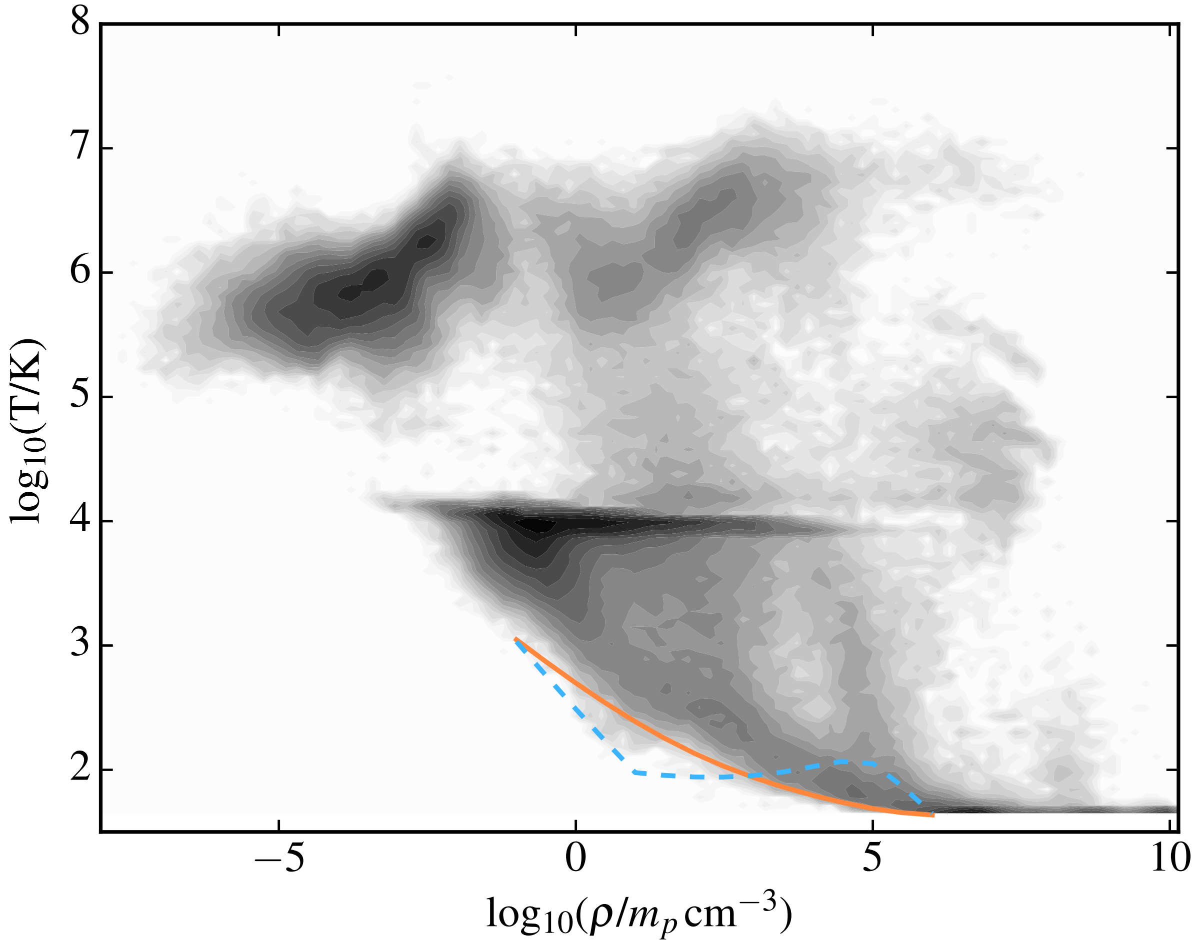

Figure 4.7 Phase diagram for the gas roughly at the final apocentre of the merger before the coalescence of the galactic cores ($t \simeq 4997$ Myr after the beginning of the simulation). The orange and dashed light blue lines delineate the gas equilibrium from Spaans & Silk (2000) for solar ($Z = Z_\odot$) and super-solar ($Z = 3Z_\odot$) metallicities respectively.

we briefly summarise their properties. The initial systems are equilibrium discs made of a Navarro-Frank-and-White dark matter halo (Navarro, Frenk & White, 1996, 1997), an exponential stellar and gaseous disc (e.g. Hernquist 1993) and a bulge following the profile proposed by Hernquist (1990). The dark matter halo has a total mass of 10^{12} $\mathrm{M}_\odot$. The disc has a scale length of 3.5 kpc and a total mass of 4×10^{10} $\mathrm{M}_\odot$, whose 10% is in gas. The bulge has a total mass of 8×10^{9} $\mathrm{M}_\odot$ and a scale radius of 0.7 kpc. With these choices the resulting galaxy model is a typical massive late-type spiral, consistent with the predictions of abundance matching at $z = 0$ (e.g. Behroozi, Wechsler & Conroy 2013; Moster, Naab & White 2013) and a gas fraction of 10% in the disc, rather typical for Sb/Sc galaxies. The gaseous disc is sampled with 10^5 particles of mass 4×10^4 $\mathrm{M}_\odot$. Both the stellar disc and the bulge are also sampled with 10^5 particles, corresponding to a particle mass of 3.6×10^5 $\mathrm{M}_\odot$ and 8×10^4 $\mathrm{M}_\odot$, respectively. The dark matter halo is sampled with 10^6 particles. Note that we use several species of dark matter particles in order to increase the

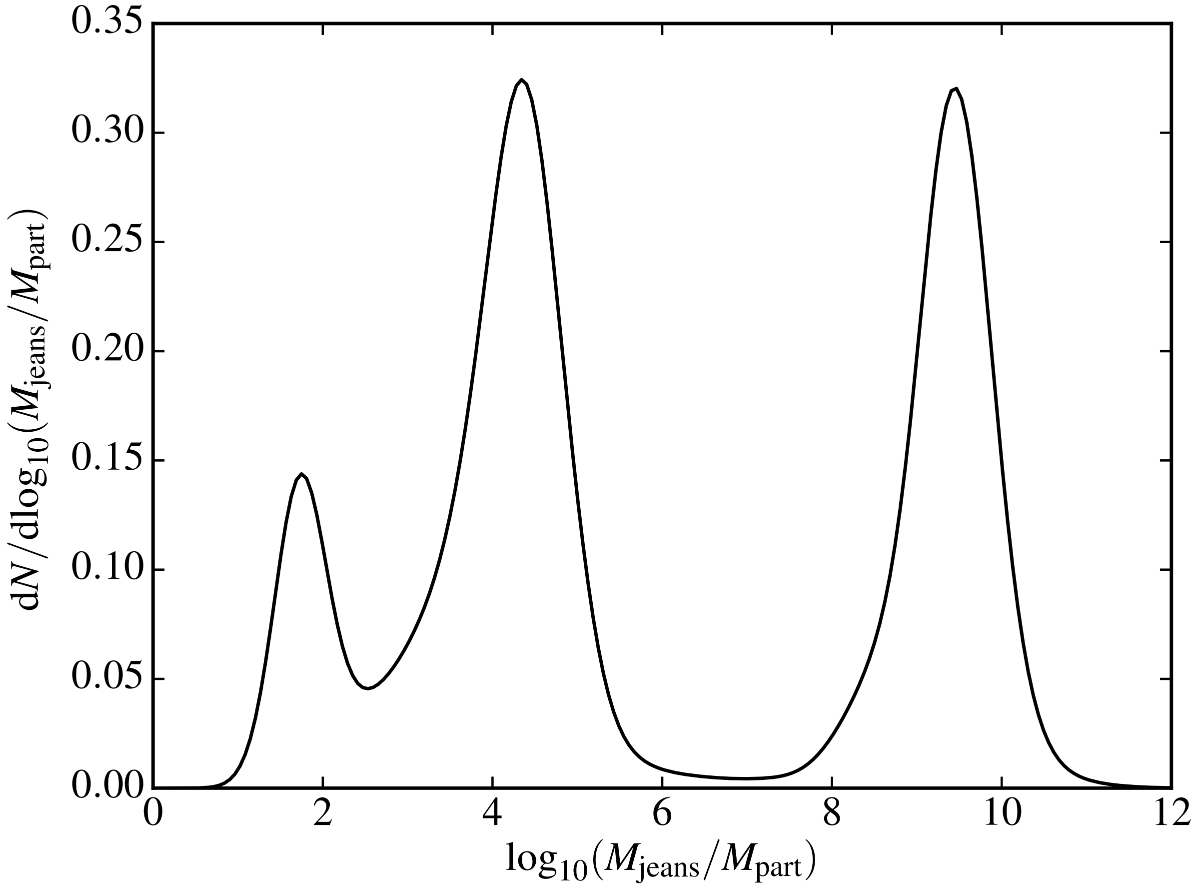

Figure 4.8 Ratio of particle mass to local Jeans mass during the final stage of the simulation. Our imposed pressure floor ensures that the local Jeans mass $M_{\rm jeans}$ is always resolved by approximately 10 particles, ensuring that any clumping we see in the simulation is physical.

effective resolution in the inner regions up to a particle mass of 8×10^4 $M_\odot$ following e.g. Kazantzidis, Magorrian & Moore (2004). The initial softening lengths are 100 pc for all particle species. Both galaxies host a supermassive black hole modelled as a collisionless particle of 2.6×10^6 $M_\odot$ initially sitting at the centre of the bulge. We have verified that the individual discs are in acceptable equilibrium by evolving them in isolation for several Gyr. The centres of the two halos are initially separated by > 500 kpc and the two systems are on an in-plane parabolic orbit with a pericentre of 50 kpc, appropriate for a cosmologically-motivated merger geometry (Khochfar & Burkert, 2006). Star formation and feedback are turned on from the beginning in order to generate a realistic multi-phase gas medium and multiple stellar populations well before the merger is completed.

We stop the simulation just before the second pericentric passage (at approximately 4.9 Gyr), when the two cores are separated by ~ 5 kpc. We define a spherical region of 35 kpc around the merger remnant for particle resampling. This is sufficient to ensure that the boundaries will not interfere with the centre on timescales of interest (~ 100 Myr), a choice that was successfully tested and utilised in a number of our previous works

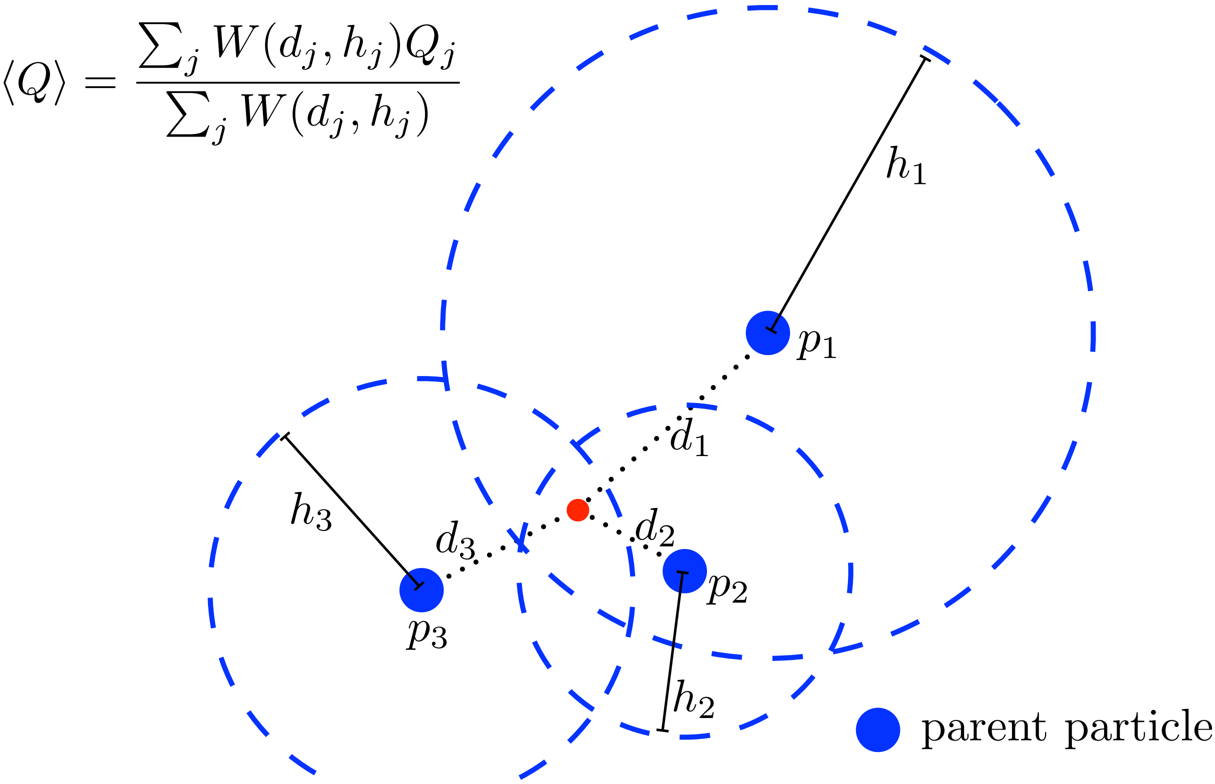

Figure 4.9 Graphical sketch of the procedure used to interpolate quantities such as temperature, metal content, etc., on to the split particles. Big blue dots p_j denotes parent particles with their smoothing length h_j (blue dashed circles) and are distant d_j from the targeted child particle, represented by the small red dot. The average in the upper-left corner reproduces equation (4.2); see the text for additional details.

(Mayer et al., 2007). We split these "parent" particles into eight "child" particles and distribute them randomly within the SPH smoothing kernel around each parent. We split only the baryonic species while the dark matter remains unsplit and provides a coarse-grained background density, as in Mayer et al. (2007). The child particles receive 1/8 of the mass of their parents and inherit the same velocity, ensuring that we conserve mass, linear momentum and angular momentum of the original parent particle. After that, we also have to assign to the child particles other simulations quantities, such as local temperature, metal content etc. In order to minimise the introduction of random noise in assigning the positions of the new particles, we do not simply assign to them the same value of the original parent particle. Instead, we first compute the local kernel of each parent particle using the positions and masses of the parent particles only. For this step, we restrict ourselves to computing the kernel radius using only 16 neighbours in order to prevent excessive blurring of boundaries in the flow. Then, for each quantity Q, a child particle is assigned the value of Q_j obtained by averaging that of each parent particle j that encompasses the same child particle within its own smoothing length h_j. The average is weighed according to the local value of the smoothing kernel W of the parent particle, namely we use:

$$\langle Q \rangle = \frac{\sum_j W(d_j, h_j) Q_j}{\sum_j W(d_j, h_j)}, \tag{4.2}$$

where the sum extends on the parent particles that encompasses the the child particle in their smoothen length h_j and d_j is the distance between the target child particle and the j-th parent particle. The procedure is briefly sketched in Figure 4.9. The softenings at this stage are set to 50 pc for the stars and gas, following the usual scaling of softening parameter ϵ with particle mass m_p, i.e. $\epsilon \propto m_p^{1/3}$. The resampled region includes 7.3×10^5 gas, 3.8×10^6 star, and 2.5×10^5 dark matter particles. The particle splitting is accomplished in part by reusing routines from SKID[4] (Stadel, 2001) to compute the densities and neighbour lists for the gather-scatter scheme.

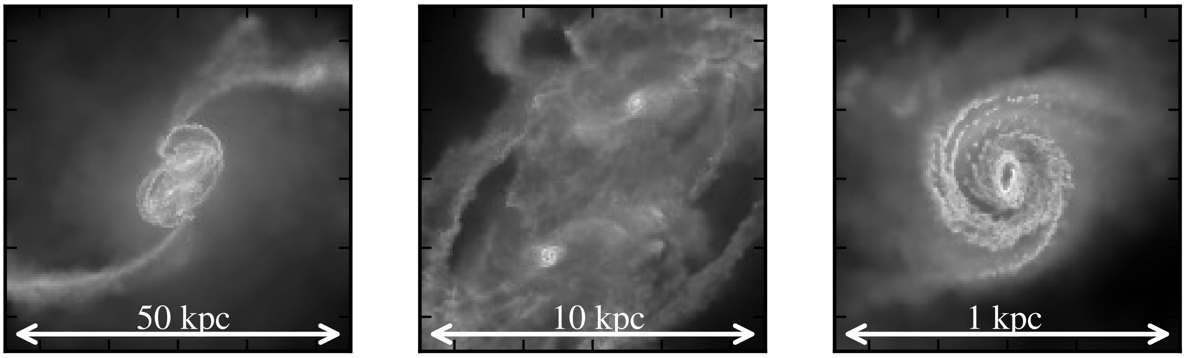

Figure 4.10 Maps of the gas density of the system just after the particle splitting procedure zooming on to the south-west galaxy core. Both galaxy cores are clearly visible and the supermassive black holes are embedded in their centres.

Figure 4.10 shows the system on several different scales just after the splitting. The left panel shows the resampled region. The two nuclear disc cores are easily identifiable in the middle panel, which also reveals the turbulent nature of the gas. The rightmost panel shows a zoom-in of one of the two nuclei, showing the dense disc around one of the supermassive black holes about 500 pc across.

During the initial stages of the merger, the supermassive black holes reside within a softening length of the potential minimum of the parent nuclear disc. However, when we switch to higher resolution, such a displacement would be much larger than the minimum separation that can now be resolved, possibly leading to artefacts in the pairing process. Therefore, at the time of splitting we also shift the supermassive black holes so that they sit at the centre of mass and match the centre of mass velocity of their respective discs. For the recentering, we determine the centre of mass and velocity of particles within a 1 kpc sphere around each supermassive black hole.

---

[4] `https://hpcforge.org/projects/skid/`

A final modification of the simulation is done at the last apocentre passage before the completion of the merger at $t = 4998$ Myr from the beginning of the simulation, when the two supermassive black holes are at a separation of ~ 500 pc and still in two separate nuclear discs with size ~ 150 pc. Here and in subsequent discussions we represent the time as $\tau = t - t_0$, where $t_0 \simeq 5001$ Myr is the last point when the cores are still distinguishable. Given this definition, the last apocentre passage corresponds to $\tau \simeq -3$ Myr. At this point we increase the force resolution to 1 pc for the gas particles and the supermassive black hole particles only in order to allow for the possibility of the supermassive black hole orbits to decay to parsec scales, in a similar fashion as in Mayer et al. (2007) and Escala et al. (2004, 2005). The stars which have already formed maintain their softening as the dark matter particles, while newly formed stars inherit the (smaller) softening of their parent gas particles. This results in a mismatch between the mass and force resolution in the stellar component. However, we expect the most important contribution to dynamical friction to come from particles closest to the supermassive black holes, i.e. those in the nuclear discs. This means that for the purposes of the orbital evolution, the newly formed stars and the gas particles will have the largest effect. These all have an identical force resolution of 1 pc. Just as in the particle splitting step, we recentre the supermassive black holes with respect to the centre of mass and bulk velocity of their surrounding gas within a distance of ~ 150 pc from each supermassive black hole. After this final modification, the simulation is evolved until the supermassive black hole orbit decays down to the scale of the softening length. We discuss the orbital decay of the supermassive black holes and the tests on the possible numerical impact of our strategy in Section 4.3.2.

4.3.2 Results

Post-merger phase

During the final passages before the completion of the merger, the two cores undergo substantial starbursts breaking up the homogeneity of the interstellar medium. These starbursts, reaching star formation rates of $\sim 80\ \mathrm{M}_\odot\,\mathrm{yr}^{-1}$, have the effect of blowing out the gas from the central region, delaying the formation of the nuclear disc for the final supermassive black hole orbit decay.

The effect of the starburst is visible in Figure 4.11, where we show a single output at several different scales, just after the two cores have fully merged at $\tau \simeq 2$ Myr. At the scales of tidal tails (several kpc), we can see that the gas structure is not smooth but has instead become clumpy and inhomogeneous due to the multiphase nature of the interstellar medium. In the inner region, the gas structure is highly irregular at this stage, without any clear evidence of an ordered rotational bulk motion. Furthermore, while the gas is funnelled into the centre, the star formation rate remains at several $\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ for several 10^7 years, which maintains a relatively high supernova activity. Figure 4.12 shows

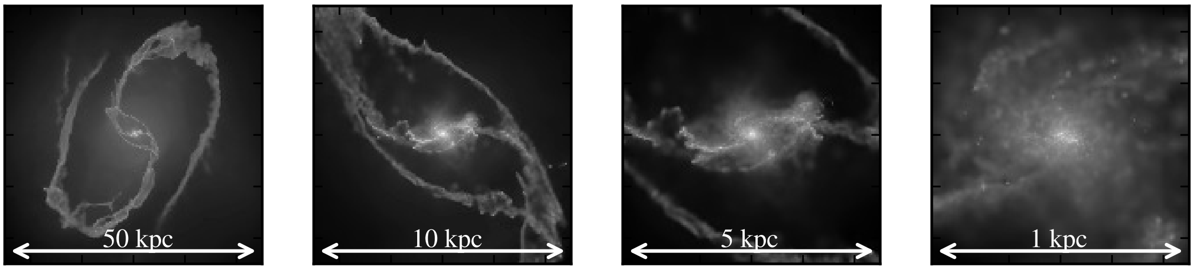

Figure 4.11 Sequence of face-on projections of the gas distribution at $t = 5003$ Myrs ($\tau \simeq 2$ Myr), just after the two cores become fully merged. The tidal tails are broken up due to feedback and multiphase interstellar medium, reducing the efficiency of accretion into the central region.

an edge-on map at $\tau \simeq 17$ Myr, i.e. after the full coalescence of the two galactic cores and the formation of a central circumnuclear discs, with the velocity field overlaid. The supernova-powered winds stir the gas surrounding the central disc, causing asymmetric outflows and inflows, as is clearly seen both above and below the central disc plane.

The combined effect of radiative cooling and feedback on the gas during the starburst is to enforce a gravitoturbulent state with widespread fragmentation in the high density gas, i.e. in the central region of the merger remnant. Fragmentation, a natural outcome in a gravitoturbulent medium (Agertz et al., 2009), leads to the formation of massive and dense clouds of cold, star-forming gas. We use the group finder SKID (Stadel, 2001) to identify groups of gravitationally bound particles in every output. We set a mean density threshold to $\rho_{\rm min} = 2$ H cm^{-3} and a linking length for the Friends-of-Friends algorithm to five times the minimum softening length, i.e. $\ell_{FF} = 5$ pc. Note that we use both, gas and stars in clump identification and $\rho_{\rm min}$ pertains to density calculated from both particle species. The density threshold ensures that we only identify the dynamically most interesting, densest particle groups. We show the resulting clump mass distributions at several representative times during the system's evolution in Figure 4.13. In particular, we selected only the clumps in the central kiloparsec of the merger remnant, most of which are actually in the inner few hundreds of parsec. The vertical dashed line indicates the mass of the supermassive black hole particles, which some of the clumps clearly exceed, in particular immediately after the galaxy cores' coalescence. At that time, many of the clumps are quite gas-rich, but due to the high densities they quickly convert much of their mass to stars. The clumps therefore quickly become akin to dense star clusters rather than gas clumps. We address the impact of these massive clumps on the supermassive black hole orbital decay below.

Disc Rebuilding Phase

The central region quickly (in a few Myr) recovers from the starburst and the circum-

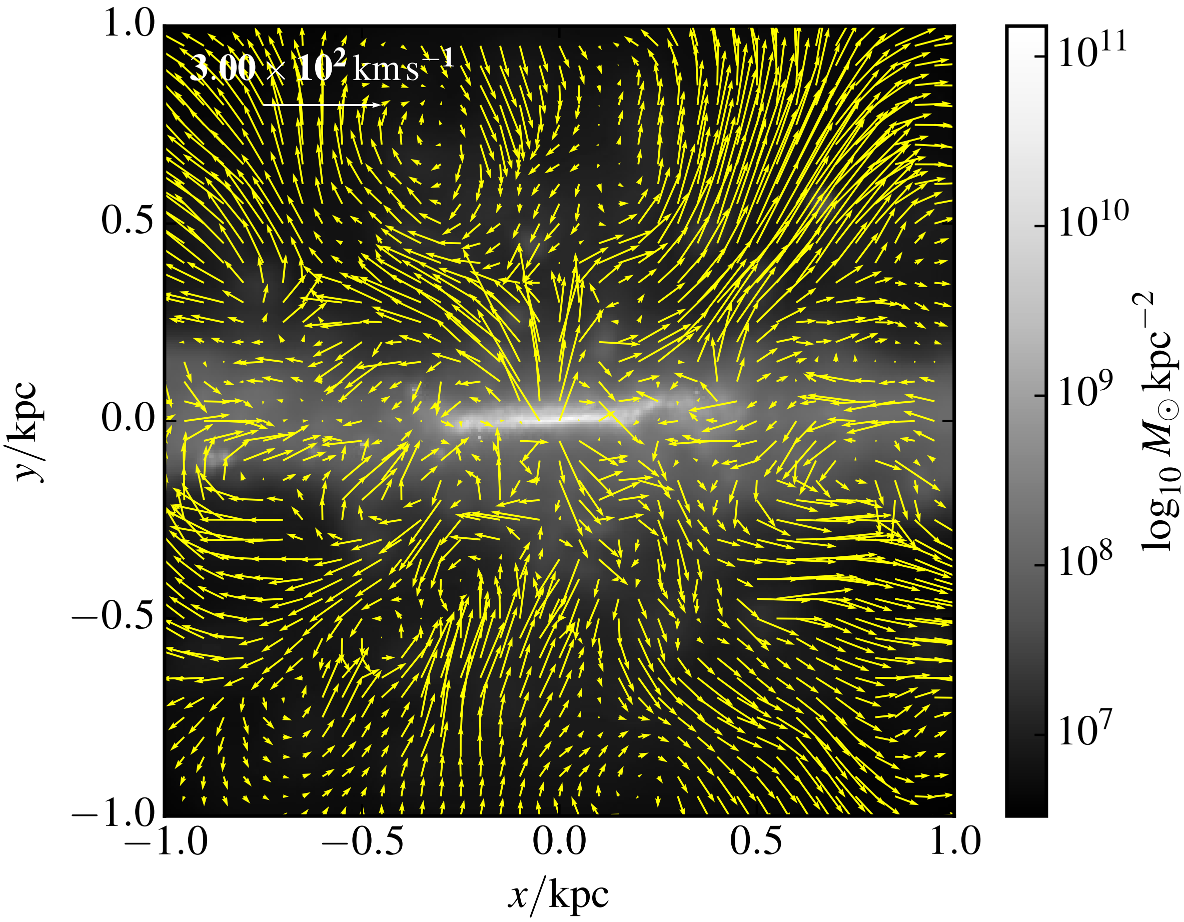

Figure 4.12 Density map after the merger at $\tau \simeq 17$ Myr with velocity field overlaid. The velocity field is generated for a thin slice, while the map is showing a projected column density. Supernova-powered outflows from the central region with velocities of several hundred km s^{-1} are not uncommon.

nuclear disc begins to reform. This is shown in the sequence projected gas-density maps of Figure 4.14. The left and right column pairs of the figure show face-on and edge-on projections respectively at different times from $\tau \simeq 0$ Myr to $\tau \simeq 6$ Myr. This sequence of image shows that after the galaxy cores' coalesce the inner region of the merger remnant recovers a disc-like morphology quickly, in $\lesssim 10$ Myr. Figure 4.15 shows the time evolution of stellar and gas density profiles in the central 100 pc (top panel) and the radial dependence of $z_{\rm rms} = \dfrac{1}{N}\sqrt{\sum_{i=0}^{N} z_i^2}$ (non-parametric proxy for thickness; bottom panel) at different times. Owing to a period of intense star formation during the merger, the stellar density here exceeds the gas density by a factor of 100-1000. This is different than the situation explored in previous studies where the nuclear gas disc dominated the mass distribution (e.g. Mayer et al. 2007). However, Dotti et al. (2007) used idealised simulations of supermassive black hole pairs in both stellar and gaseous nuclear discs to show that the stellar (or gaseous) fraction affects the dynamics only marginally, rather is the total surface density of baryons in the disc mid-plane that sets the magnitude of

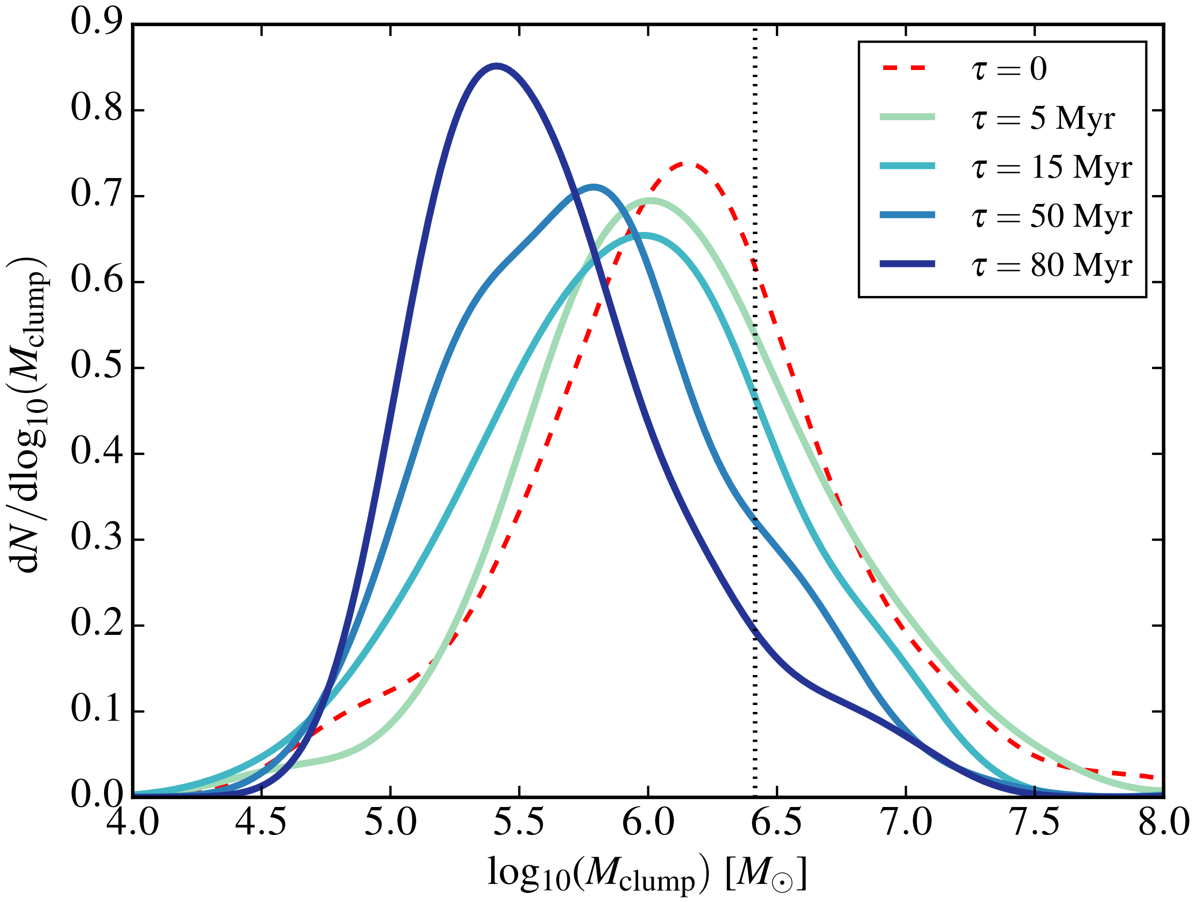

Figure 4.13 Clump mass distributions per logarithmic mass interval at five different times after the galaxy cores' coalescence as marked in the legend. The distributions are only for clumps in the central kiloparsec. The vertical line marks the mass of the supermassive black holes.

the drag onto the supermassive black holes. Furthermore, while a nuclear disc does form, the accretion is hindered by the fact that within 0.5 kpc of the centre, $\sim 5 - 10\%$ of the mass is locked up in massive clumps (this fraction is highest soon after the merger and decreases at later times). As can be seen from Figure 4.13, the clumps form even at late times (although their masses decrease somewhat), continuously stirring the disc. In most cases, the dense gas clumps at early times convert most of their mass into stars but remain gravitationally bound and only slowly get disrupted. Note that due to the short timescales in question ($\lesssim 10$ Myr) to form the clumps and convert them into stars, the supernova feedback has little effect on regulating the clump masses.

The gaseous and stellar discs are significantly flared, as shown by the strong evolution of z_{rms} (model-independent proxy for scale height) as a function of radius at all times. While the thickness of the stellar component remains largely fixed throughout the 80 Myr of evolution, however, the decrease in the thickness of the gas component is evident especially in the inner regions. This is due to the accumulation of gas in the central region as the disc reforms and the chaotic merger-induced disc structure settles down.

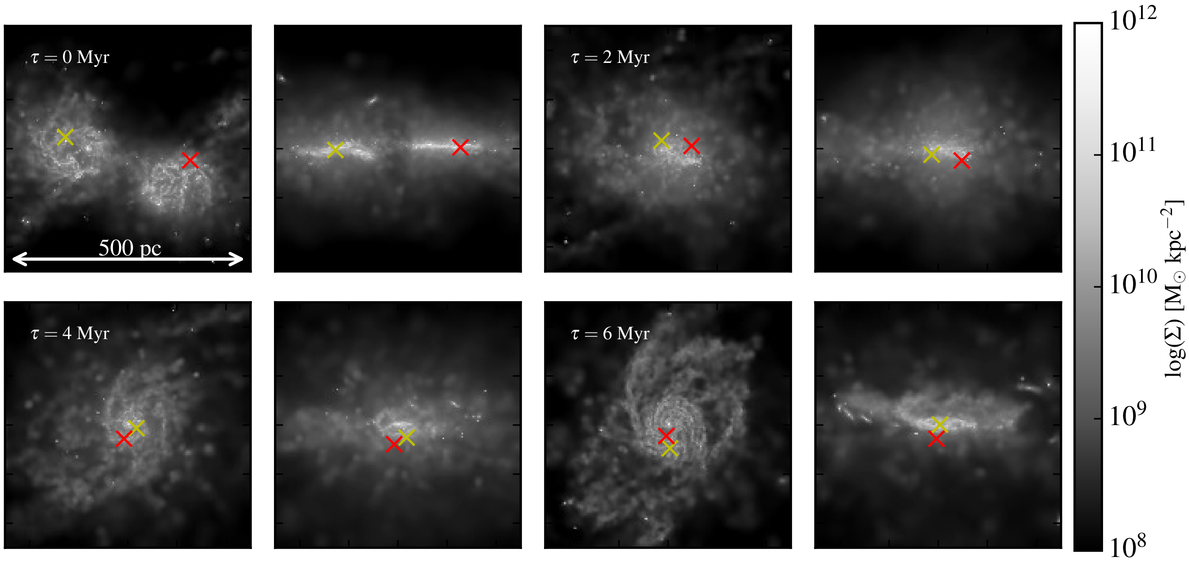

Figure 4.14 Face-on and edge-on gas density projections at several times during the final phase of the merger, after the last apocentric passage. The two supermassive black hole particles are indicated by yellow and red crosses.

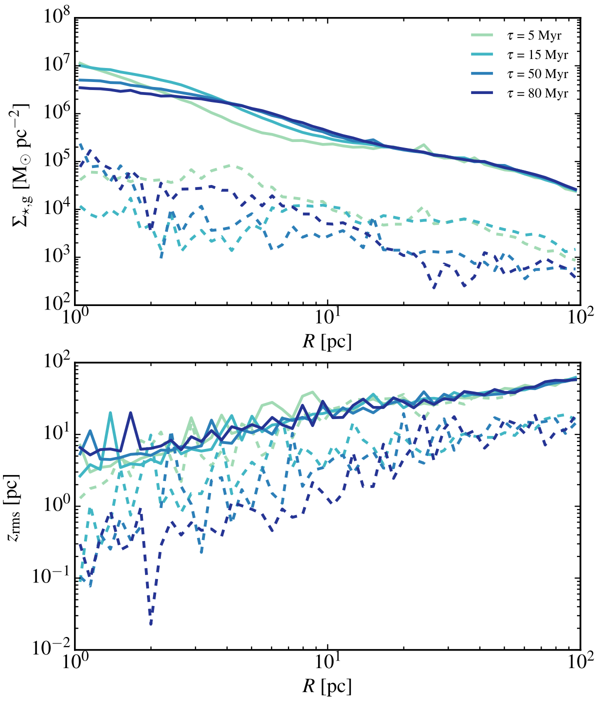

Figure 4.15 Surface density and $z_{\rm rms}$ (see definition in the text) as a function of radius (top and bottom panels respectively) for stars (solid) and gas (dashed) at four representative times, as indicated in the legend.

Figure 4.16 shows the ratio between rotational velocity v and 3D velocity dispersion σ of the gas disc, quantitatively demonstrating the disc rebuilding phase. Immediately after the merger ($\tau = 0$ and $\tau = 2$), $v/\sigma < 2$, indicating lack of rotational support and

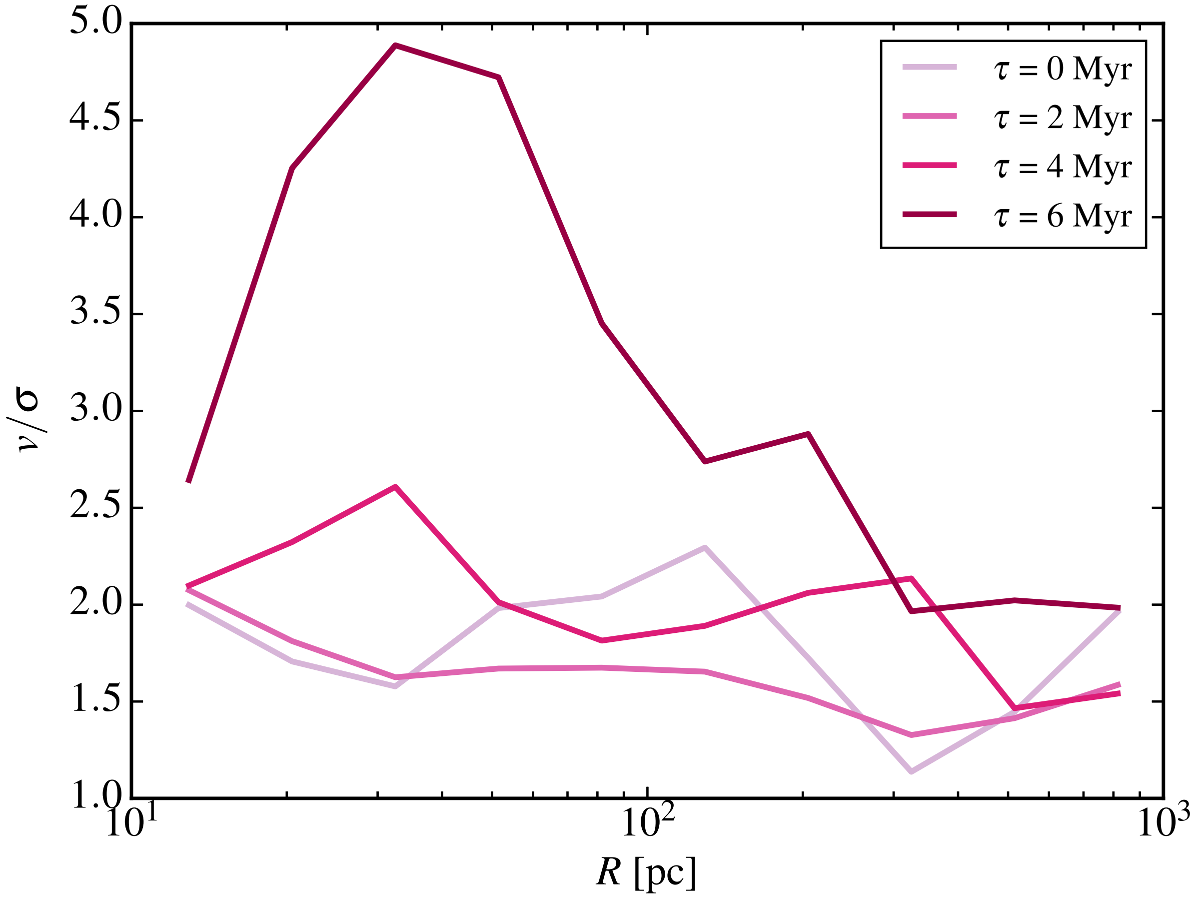

Figure 4.16 The ratio of tangential velocity v to the 3D velocity dispersion σ at several times shortly after the merger is complete. The disc rebuilds on a timescale $\lesssim 10$ Myr, evidenced by the rapid rise in v/σ, and extends out to ~ 200 pc.

ordered bulk motion throughout the central region. However, after just 4 Myr $v/\sigma > 2$ in the interior 300 pc and by $\lesssim 10$ Myr after the merger is complete, the inner several hundred parsecs contain a rotationally-supported, kinematically cold gas disc, confirming the visual impression from the morphology shown by Figure 4.14. Nevertheless, the gaseous disc remains much less massive than the stellar component (Figure 4.15).

Figure 4.16 shows the most "optimistic" measure of v/σ, which is calculated simply as the ratio of the average tangential velocity component and the velocity dispersion in each radial bin. However, observationally, v/σ is typically measured from the line-of-sight velocity and the associated velocity dispersion based on spectral line shift and width respectively. We crudely model this type of v/σ determination in Figure 4.17, showing $v_{\rm los}/\sigma$ for a range of inclination angles (0° corresponds to face-on) at $\tau \simeq 6$ Myr. The majority of $v_{\rm los}/\sigma$ values lie in the range of 1-2, which could be interpreted as a thickened, highly turbulent disc.

In a recent study targeting 17 local (ultra) luminous infrared galaxies, Medling et al. (2014) found that in most cases (16 out of 18 nuclei) they contain a dense nuclear disc

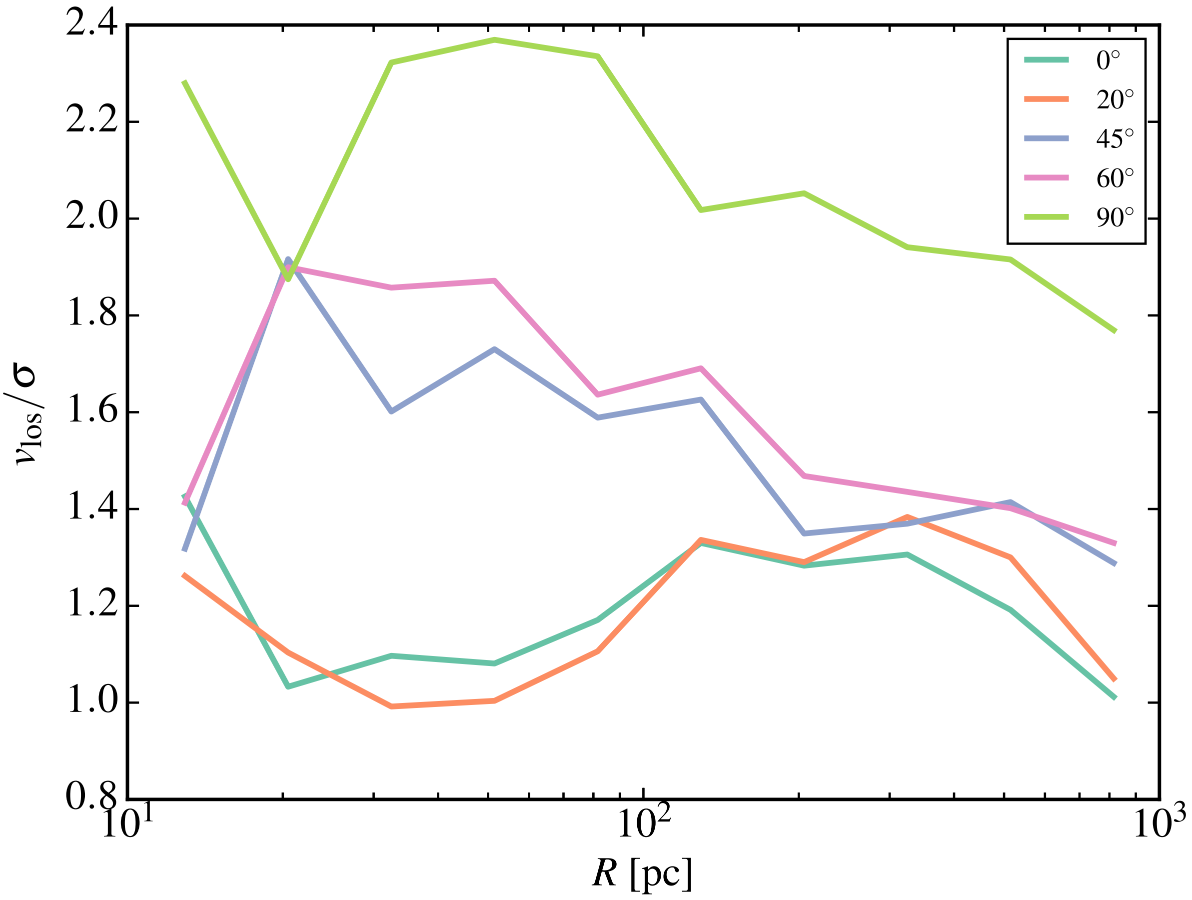

Figure 4.17 $v_{\rm los}/\sigma$, where $v_{\rm los}$ is the mean velocity along the line of sight. Colours represent different inclinations of the central disc. The output shown is at $\tau = 6$ ($t = 5007$ Myr).

composed of gas and stars. The observed discs mostly have effective radii of a few hundred parsecs and dynamical masses estimated in the range of $10^8 - 10^9$ $M_\odot$. The stellar populations appear young (ages < 10 Myr), implying that they are associated with in-situ star formation in the nuclear discs themselves, analogous to the disc in our simulations. Furthermore, they find that v/σ in their sample ranges from 1-2, in good agreement with the values in Figure 4.17. We can therefore be reasonably confident that we are capturing some of the essential processes of the nuclear disc rebuilding.

Supermassive black hole orbital decay

In the previous section we highlighted some of the most important aspects of the morphological evolution of the merger remnant. A prominent feature of the merger in our simulation is that although in the initial conditions the system lies in a single plane, this symmetry is broken during the final stages of the merger (Figure 4.14). As discussed above, the starburst responsible for the symmetry breaking is accompanied by the formation of massive clumps. The combination of global non-axisymmetric disc torques and gravitational encounters with massive clumps significantly perturbs the dynamics of the two supermassive black holes.

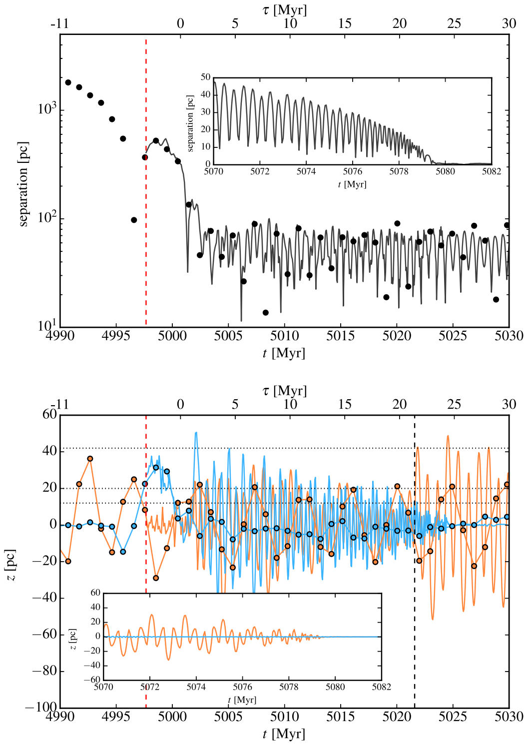

Figure 4.18 Upper panel: separation of the supermassive black holes as a function of time starting at $\tau = -11$ Myr. The black continuous line shows the result of our simulation, while the grey circles show a twin run where no softening reduction at $\tau \simeq -3$ Myr is performed. Bottom panel: motion perpendicular to the disc plane for each supermassive black hole. Orange and light blue continuous lines show the result of our simulation. Orange and light blue circles interconnected by continuous lines show the same quantity for a twin run where no softening reduction at $\tau \simeq -3$ Myr is performed. From top to bottom, the dotted horizontal lines show the $z_{\rm rms}$ of stars at 50, 10, and 5 pc, while the black vertical dashed line at $\tau \simeq 20$ Myr marks the (orange) supermassive black hole-clump interaction described in the text. In both panels, the insets show the last $\sim$ 10 Myr of evolution before supermassive black hole coalescence for the main run discussed in the text, while the red vertical dashed line at $\tau \simeq -3$ Myr marks the softening reduction.

Figure 4.18 summarises the time evolution of the supermassive black hole pair roughly from the last apocentre to the final coalescence of the two supermassive black holes (i.e. when their separation reaches the force resolution). In particular the upper panel shows the time evolution of the separation between the supermassive black holes, whereas the lower panel shows their motion in the z direction (i.e. the axis perpendicular to the reformed circumnuclear disc). The vertical red dashed line in both panel marks the point at which we decrease the softening to 1 pc. The grey points in the upper panel as well as the orange and light blue points with lines in the lower panel show the separation and the vertical motion, respectively, in a twin run where no softening alteration (i.e. no softening reduction at $\tau \simeq -3$ Myr) was performed. The upper panel shows that the reduction of the softening has no obvious repercussions on the separation of the two supermassive black holes immediately for the initial 10 Myr. Remarkably, the lower panel shows also that the lower resolution resolution run (light blue dots with line) and its high-resolution counterpart (light blue line only) do not show significant deviations just after the softening change as both supermassive black holes sink toward the disc mid-plane in the following ~ 3 Myr. The other supermassive black hole particle in the high resolution run (orange line only) differs from its counterpart (orange dots with line) due to the recentering that is applied at the same time as the softening change. Nonetheless, these findings indicate that our numerical strategy for resampling does not introduce artefacts in the evolution of the system. Moreover, the low resolution simulations show a qualitatively similar behaviour and a similar orbital decay timescale relative to the higher resolution runs up to a binary separation that they can resolve (about a couple of softening lengths).

Although the initial conditions were the same, the evolution of the supermassive black hole pair inside the circumnuclear disc differs dramatically from that in the simulations presented in Mayer et al. (2007). Indeed, Mayer et al. (2007) found that the supermassive black holes reached a separation of order the resolution limit of 2 pc $\sim 5-10$ Myr after the last apocentre, while in our simulation the supermassive black holes stall at a separation of several tens of parsec for ~ 80 Myr before sinking further. Such a qualitatively different behaviour is caused by the different density structure and thermodynamical conditions of the circumnuclear disc that provides the drag onto the supermassive black holes. In Mayer et al. (2007) simulations, after the final merger of the gaseous cores, the use of an effective equation of state (and therefore the lack of the final starburst and the consequently smoother gas distribution) favours the rapid formation of a dense, smooth, gas-dominated circumnuclear disc containing the supermassive black holes, which enables and accelerates their coalescence (see also Chapon, Mayer & Teyssier 2013). On the other hand, the aftermath of the merger in our simulation destroys the plane-symmetry of the system and excites the out-of-plane oscillations of the two supermassive black holes (at $\tau \gtrsim 0$ Myr). Because the densest component (stars and gas) is distributed in a disc, the increased vertical oscillations result in a reduced dynamical friction when the supermassive black holes move in the diffuse envelope of the disc, and therefore the supermassive black hole orbital decay timescale is prolonged, as found by Fiacconi et al. (2013, see Section 4.2). The decay of the binary, of the binary, especially as measured by the vertical oscillation amplitude, continues until at $\tau \sim 25$ Myr one of the supermassive black holes (light blue) begins to orbit within the misaligned nuclear disc. As it approaches the centre, it encounters a weak stellar bar which measures a few parsec across. The supermassive black hole particle orbit is strongly torqued by the small bar and it loses all of its angular momentum very rapidly, becoming pinned to the potential minimum at the centre.

At the same time, the vertical oscillation amplitude of the other supermassive black hole particle (orange) also decreases slowly between $\tau \simeq 0$ Myr and $\tau \simeq 20$ Myr. However,

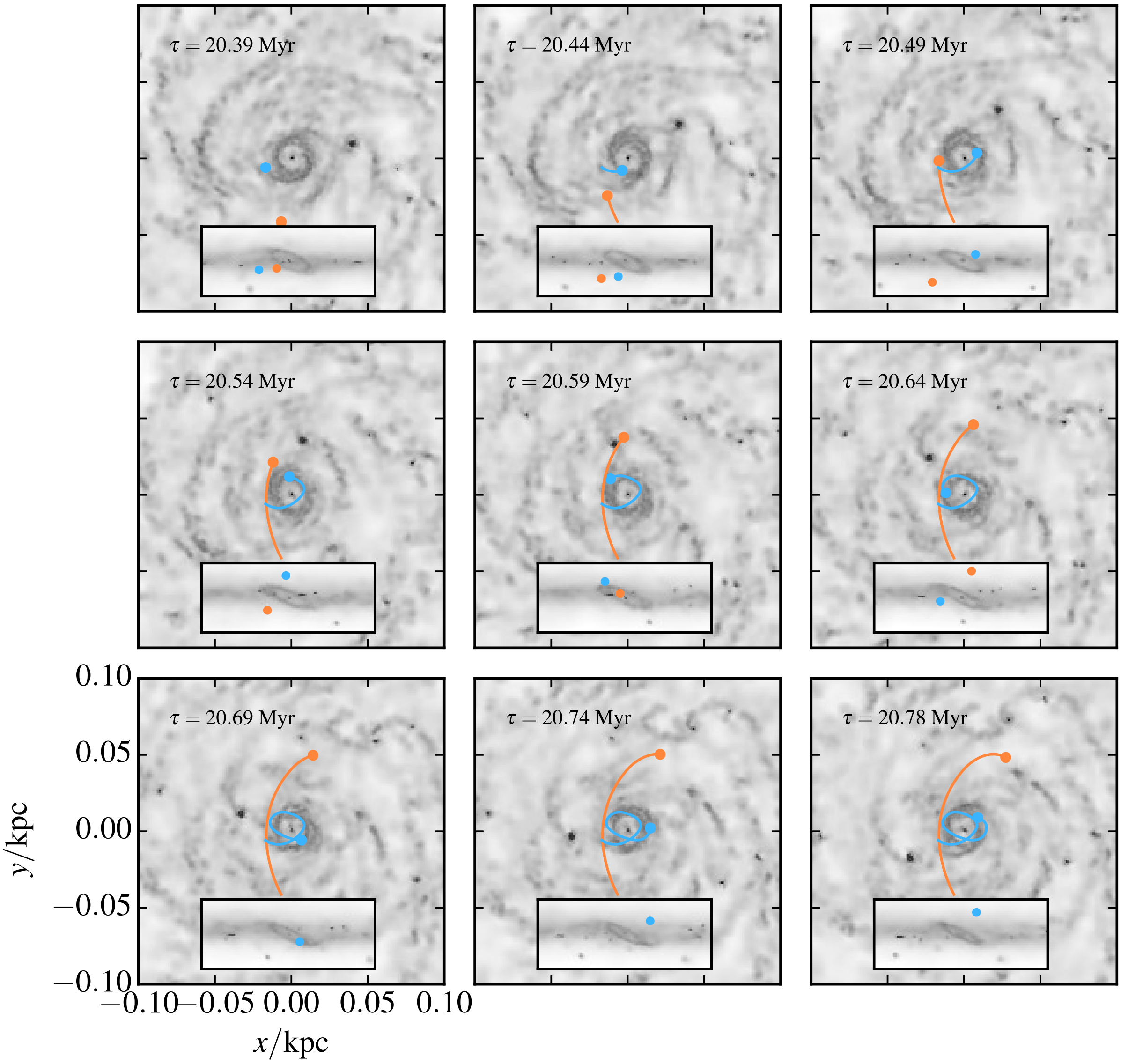

Figure 4.19 Sequence of face-on gas density projection showing the close encounter between the (orange) supermassive black hole and a massive clump. Each panel shows the trajectory of both supermassive black holes in the face-on view, while the edge-on view of each time is shown in the insets The middle panel shows the closest approach and is marked in Figure 4.18 by a vertical dashed black line at $\tau \simeq 20$ Myr.

just after $\tau \simeq 20$ Myr, the (orange) supermassive black hole experiences a close interaction with a massive clump, which significantly perturbs its orbit. This interaction is shown in Figure 4.19 as a sequence of gas density projections with the positions of both supermassive black hole particles superimposed. A massive clump with mass $\sim 2.4 \times 10^7$ $\mathrm{M}_\odot$ orbits at $\sim 30 - 40$ pc from the centre of the circumnuclear disc (recall that the mass of the supermassive black holes is $2.6 \times 10^6 \mathrm{M}_\odot$). At $\tau \simeq 20.55$ Myr, the (orange) supermassive black hole passes at a few parsec from the massive clump (the close passage is confirmed by both the face-on and edge-on views in the figure). After that interaction, the amplitude of its vertical oscillation increases again up to ~ 50 pc and then it spends most of is orbital time outside the disc plane, where the density background, composed by both stars and gas, is lower than in the disc plane. Therefore, the black hole feels a smaller drag on its orbit due to dynamical friction. We have estimated the dynamical friction timescale due to only the background density (using calculations similar to those in Fiacconi et al. 2013 and described in Section 4.2) in the inner bulge to be ~ 50 Myr, roughly consistent with the final decay timescale. Note that the supermassive black hole is moving supersonically through the background outside the disc plane at typical mach numbers of 2-3. In principle this should amplify the drag caused by the gaseous component (Ostriker, 1999; Chapon, Mayer & Teyssier, 2013), but its density is sufficiently low that the resulting dynamical friction does not dominate the evolution.

The supermassive black holes reach the minimum resolved separation at $\tau \sim 80$ Myr as shown in the insets of Figure 4.18. During the last few Myr of the decay, the orbit of the (orange) supermassive black hole has an eccentricity $e = (r_\mathrm{a} - r_\mathrm{p})/(r_\mathrm{a} + r_\mathrm{p}) \simeq 0.7$ (where r_a and r_p are the apocentre and pericentre radii, respectively) and just before the binary decays to our resolution limit $e \sim 0.4$. This is consistent with the findings of Section 4.2 (Fiacconi et al., 2013): the orbits of supermassive black holes within a clumpy interstellar medium remain eccentric down to the resolution limit. The fact that the binary might approach the gravitational wave regime before circularising can be critical for the subsequent gravitational wave emission (Mayer, 2013).

4.4 Discussion and conclusions

We probed the dynamics of massive black hole pairs in clumpy and inhomogeneous environments. First, we adopted idealised simulations of gaseous circumnuclear discs, comparing the orbital decay below ~ 100 pc in smooth and highly inhomogeneous discs. Then, we employed a self-consistent simulation of a merger between two disc galaxies hosting a supermassive black hole at their centres to study the evolution of the pair from > 10 kpc down to 1 pc separation, where the interstellar medium is naturally shaped by the interplay of radiative cooling, star formation and stellar feedback. We now discuss our results in a broader scope.

4.4.1 Dynamics and orbital decay timescale

Massive black hole pairs in clumpy circumnuclear disc can experience different processes that alter the orbital decay timescale when compared to equivalent smooth conditions. The black holes can gravitationally interact with the spiral (material) arms that develops in the circumnuclear disc because of the disruption of massive clouds and global instabilities. Spiral arms transfer angular momentum outward to the incoming black hole from larger radii, behaving effectively as non-steady angular momentum barriers that prevent the black holes from loosing angular momentum and from spiralling in. Massive clouds can gravitationally interact with the black holes in two-body interactions. When the cloud is less massive than the black hole, the latter can disrupt the cloud and accrete its mass. The black hole effective mass then increases and it allows for a faster decay, sometime as fast as a few millions of years. However, when the cloud mass is comparable or larger than the black hole mass, the interaction can lead to gravitational slingshots. Gravitational slingshots frequently perturb the orbit of the moving black hole, stochastically increasing or decreasing the separation of the pair. They can even eject the black holes from the disc plane, having a net retarding effect on their orbital decay, since in this case the dominant source of drag on the black hole is the lower density stellar background. Overall, a clumpy interstellar medium can stochastically enlarge the range of orbital decay timescales from scales ~ 100 pc down to the formation of a bound binary. This conclusion is supported by both the idealised simulations of the circumnuclear discs and the self-consistent merger simulation.

When a black hole is ejected from the gaseous disc, we can estimate its orbital decay timescale $\tau_{\rm decay}$ assuming it is in circular motion inside a Plummer bulge of total mass $M_{\rm b}$ and scale radius b consistent with the values adopted in Section 4.2. By using standard dynamical friction (Chandrasekhar, 1943),we determine $\tau_{\rm decay} \simeq \tau_{\rm decay}^{\rm (ext)} + \tau_{\rm decay}^{\rm (int)}$, where:

$$\tau_{\rm decay}^{\rm (ext)} \simeq \frac{0.5305}{\ln\Lambda} \left(\frac{M_{\rm b}}{M_\bullet}\right) \left[(r_{\rm in}/b)^{7/2} - 1\right] t_{\rm dyn}, \tag{4.3}$$

is the time to move from an initial radius $r_{\rm in}$ to $\sim b$, and:

$$\tau_{\rm decay}^{\rm (int)} \simeq \frac{0.2172}{\ln\Lambda} \left(\frac{M_{\rm b}}{M_\bullet}\right) t_{\rm dyn}, \tag{4.4}$$

is the decay time inside the core. As usual, $\ln\Lambda$ is the Coulomb logarithm and $t_{\rm dyn} = 1/\sqrt{G\rho_0}$ is the dynamical time associated with central density ρ_0 of the Plummer profile. We estimate that $\tau_{\rm decay}$ would vary typically between ~ 20 and ~ 50 Myr depending on the secondary black hole mass and the properties of the stellar spheroid that we assume for our simulations. We used $\ln\Lambda \sim 5$ and $r_{\rm in} = 60$ pc. We also checked that the decay timescale would vary in the range $\sim 20 - 80$ Myr if we adopted a (de-projected) Sérsic profile for the bulge with structural parameters (mass, effective radius, Sérsic index) ranging from those typical of massive classical bulges (e.g., Andromeda galaxy, Widrow, Perrett & Suyu 2003; Courteau et al. 2011) to those of pseudobulges in late-type spirals (Fisher & Drory 2008; Fisher, Drory & Fabricius 2009). This findings are corroborated by the results of the merger simulation, where the two supermassive black holes form a loose binary after the circumnuclear disc has formed, but the violent nature of the late stages of the merger causes their orbits to have a considerable motion perpendicular to the disc plane. This delays the complete decay of the supermassive black hole pair which takes approximately 80 Myr, almost two orders of magnitude longer than previously found in analogous, but smooth, systems (Mayer et al., 2007).

4.4.2 A threshold mass for stochastic dynamical effects

We estimate a threshold mass $\mathcal{M}_\bullet$ below which a $M_\bullet$ black hole orbiting in a disc with mass $M_{\rm g}$ and scale radius R will likely be scattered by clumps with $M_{\rm cl} > M_\bullet$. The black hole and the clumps both migrate toward the centre on a timescale $\tau_\bullet$ and $\tau_{\rm cl}$, respectively. We can envision that the black hole moves in an almost steady-state environment in which massive clumps that dissolves close to the centre are continuously replaced by new clumps on the same timescale, so that the number and the spatial distribution of massive clumps remains almost constant in time, on average. This is a reasonable assumption when we focus on massive clumps only since they form because of the merging of smaller clumps during their migration on the timescale $\tau_{\rm cl}$ and dissolve when they reach the centre on the same timescale. Then, the black hole moves radially with an effective relative velocity $v_\bullet \sim R/\tau_\bullet \sim 10$ km s^{-1} with respect to the clumps. Under these crude assumptions, we can imagine that the black hole moves radially by a random walk from $\sim R$ to the centre of the disc. The condition for the scattering by clumps to be sizeable reads:

$$1 \ll N \sim R^2 n_{\rm cl}^2 \sigma_{\rm cl}^2, \tag{4.5}$$

where $n_{\rm cl} \sim N_{\rm cl}/V_{\rm disc}$ is the number density of clumps in the disc volume $V_{\rm disc} \sim h\, R^3$ with aspect ratio h, and $\sigma_{\rm cl}$ is the cross section for a gravitational interaction with a clump. Massive clumps have a size ∼5-10 pc, whereas the influence radius for interaction with the black hole $r_{\rm g} \sim G M_{\rm cl}/v_\bullet^2 \gtrsim 10$ pc for $M_{\rm cl} \gtrsim 10^6$ M$_\odot$, therefore we can estimate $\sigma_{\rm cl} \sim \pi r_{\rm g}^2$. If we assume that $\tau_\bullet$ scales as for dynamical friction, $\tau \sim (\Theta/\ln\Lambda)\,(M_{\rm g}/M_\bullet)\, t_{\rm dyn}$, where $\Theta \sim 1$, we can rewrite equation (4.5) in terms of the black hole mass $\mathcal{M}_\bullet$ below which scatterings are relevant:

$$M_\bullet \ll \mathcal{M}_\bullet \sim \frac{\left(N_{\rm cl}\, h^{-1}\right)^{1/4}}{\ln\Lambda} \left(\frac{M_{\rm cl}}{M_{\rm g}}\right)^{1/2} M_{\rm g}. \tag{4.6}$$

In gas-rich system, the maximum clump mass allowed by Toomre instability is $M_{\rm cl} \sim \eta^2 M_{\rm g}$, where η is the gas fraction of the system (Escala & Larson, 2008). Plugging that in equation (4.6), we finally obtain:

$$\mathcal{M}_\bullet \sim 7\times 10^7 \left(\frac{N_{\rm cl}}{4}\right)^{1/4} \left(\frac{\eta}{0.4}\right) \left(\frac{M_{\rm g}}{10^9\ {\rm M}_\odot}\right) {\rm M}_\odot, \tag{4.7}$$

where we assumed $h = 0.2$ and $\ln\Lambda = 5$.

Although it is only an order of magnitude estimate, this result is fairly consistent with our findings from both the idealised circumnuclear disc simulations and the merger simulation, and possibly even with a recent discovery of a supermassive black hole pair by Fabbiano et al. (2011). The threshold mass roughly corresponds to a bulge mass $\lesssim 5\times 10^9$ M$_\odot$, according to the scaling relation between supermassive black hole and bulge masses (Sani et al., 2011). The combination of these factors suggests that supermassive black hole pairs might be possibly more easily detected in late-type spirals (Sc-Sd), which are relatively gas-rich even at $z = 0$ and have stellar masses $\lesssim 10^{10}$ M$_\odot$, although the occupation fraction of supermassive black holes at such low masses is still unclear (Greene et al., 2010; Reines et al., 2011; Reines, Greene & Geha, 2013). However, these considerations neglect the details of the dual AGN activity required for the supermassive black holes to be both detected. Just after the merger phase, obscuration might be relevant or the central gas might be depleted after the starburst and mostly accumulated in to massive clumps, reducing the time during which both supermassive black holes could be active at the same time.

4.4.3 Black hole pairs and high-redshift galaxies

The simulations presented here use models that have been tailored to reproduce the conditions in low-redshift late-type galaxies. This was motivated primarily by the need to compare with previous work using similar initial conditions (e.g. Mayer et al. 2007). However, since galaxy mergers are much more frequent at higher redshift, future work will have to explore mergers between galaxies whose properties are more akin to high-z galaxies. Qualitatively, we can expect that the stochastic orbital decay regime found here will be even more relevant because at $z > 1$ galaxies with stellar masses comparable to the Milky Way appear to be more gas-rich as well as clumpier, even when they are not in a merging phase (Genzel et al., 2006; Elmegreen et al., 2009). Clumps observed in high-z galaxies are also much more massive, up to a few 10^8 M$_\odot$ (but more extended), which could imply stronger gravitational forcing of the massive black holes during encounters.

They might also affect comparatively more massive black holes, in the range $10^7 - 10^8$ M$_\odot$, as deduced from equation (4.7) when we consider disc-like systems with gaseous mass $\gtrsim 10^{10}$ M$_\odot$. In fact, this might be very relevant for black hole pairs at $z \gtrsim 2$ that will be probed by space-born gravitational wave detectors such as eLISA (Amaro-Seoane et al., 2013). Moreover, gas-rich, clumpy, massive disc galaxies at $z \gtrsim 2$ with relatively massive central supermassive black holes are likely the progenitors of massive elliptical/S0 galaxies today. Therefore, our finding may be important to understand the dynamical evolution and mass growth of the most massive supermassive black holes found in the current Universe.

4.4.4 Caveats

Recently, del Valle et al. (2015) have questioned our results, suggesting that the stochasticity of the orbital decay might be artificially enhanced by intrinsic limitations of the simulations and therefore not so frequent. They conclude that the typical orbital decay timescale in circumnuclear discs is $\sim 5 - 10$ Myr, weakly dependent on the star formation properties. While it is certainly true that simulations have limitations that might influence the final result and that we did not run a sample of simulations large enough to exclude that very long/short orbital decay timescales occupy only low probability tails of an unknown, underlying distribution, our simulations clearly show that the results are mostly dependent not only on the presence of massive clumps, but also on the orbital properties of the black hole. Indeed, the episodes of delayed orbital decay due to clump-black hole interactions are typically associated with eccentric orbits, while del Valle et al. (2015) only probed nearly circular orbits. It is not difficult to guess why little effects from gravitational interactions has been reported: when both the gas and the black holes are in circular orbits, the relative velocity of a close interaction is going to be low, favouring the tidal heating of the cloud by the black hole and the subsequent accretion. On the other hand, a black hole on an eccentric orbit may more likely interact with a clump at higher relative velocity, inducing encounters able to vary the directions of the black hole motion, redistributing the total velocity among different vectorial components. Moreover, as shown in Figure 4.3, the initial role of dynamical friction during circularisation is to reduce the orbital energy at nearly constant angular momentum. The black hole starts to loose angular momentum only after circularisation. When it moves on an eccentric orbit within an inhomogeneous environment, the structure of the interstellar medium can destroy the coherence of the trailing wake required to circularise the orbit, indirectly delaying the orbital decay process. As a final note, del Valle et al. (2015) explore the effects of star formation parameters that do not control the amount of lumpiness, but only the relative amount of stars and gas in overdense regions, which is of negligible importance in gravitational interactions between clumps and black holes.

Finally, an important caveat is that our simulations do not include the effects of black hole accretion and the associated feedback. For the merger simulation described in Section 4.3, we perform a crude estimate of the expected accretion rate assuming spherical Bondi accretion and taking into account the relative velocity between the black hole and the background:

$$\dot{M} = \frac{4\pi G^2 M_\bullet^2 \rho}{(v_{\rm rel}^2 + c_{\rm s,turb}^2)^{3/2}}, \tag{4.8}$$

where $M_\bullet$ is the mass of the supermassive black hole, $v_{\rm rel}$ is the relative velocity between the supermassive black hole and the background, $c_{\rm s,turb} = (c_s^2 + \sigma^2/3)^{1/2}$ is the turbulent sound speed with σ the gas velocity dispersion and G is the gravitational constant. Insofar as the supermassive black hole particles can be considered as test particles sampling the potential in the inner part of the galaxy, we approximate $v_{\rm rel}$ as the velocity dispersion of the stellar component which during the time shortly after the merger is ~ 200 km s^{-1}. The resulting Bondi rates are $\sim 0.05 - 0.1$ M$_\odot$ yr^{-1}. This means the black holes could double their mass during the ~ 100 Myr of decay phase in the nuclear disc, which would have only a marginal effect on dynamical friction as well as on the exchange of energy and angular momentum via torquing. Using these approximations and assuming that the accretion luminosity is given by $L = \epsilon_r \dot{M} c^2$, where ϵ_r is $\sim 10\%$, we estimate that, for the typical gas densities encountered by the supermassive black holes (recall that the supermassive black holes oscillate significantly above and below the disc plane), the expected luminosities would be $< 0.1 L_{\rm edd}$, i.e. $\sim 10^{43}$ erg s^{-1}. Only a fraction of this accretion luminosity would couple to the interstellar medium. While the coupling mechanism and efficiency is largely unknown, simulations that attempt to reproduce the observed correlations between supermassive black hole masses and host galaxy properties require a coupling efficiency $\epsilon_{\rm fb} \sim 0.005 - 0.05$, where the wide range of values is explained by the dependence on the specific thermodynamical model of the interstellar medium that simulations adopt (see e.g. Springel, Di Matteo & Hernquist 2005; Callegari et al. 2009). Therefore, we can assume that the luminosity actually coupled with the gas would be $\lesssim 10^{42}$ erg s^{-1}. We find that this amount of energy is comparable to the energy released by supernovae during the phase of supermassive black hole decay in the nuclear disc Overall, these calculations suggest that the impact of black hole feedback in the supermassive black hole decay phase would be modest since it would likely deposit at most the same amount of energy as supernova feedback, though there are huge uncertainties on the various feedback processes in the first place.

4.4.5 Concluding remarks

We have shown that the evolution of massive black hole pairs can be substantially influenced by inhomogeneities and massive clouds in the interstellar medium. We show that by using progressively more realistic simulations and we speculate on the possibility that similar or even stronger effect might be in place also at higher redshift (first confirmations of such statements are in development; Taburello et al., in preparation). The fundamental result is that the range of orbital decay time scale is enlarged toward both faster and slower value by up to an order of magnitude compared to previous calculations. Although our results do not change the overall scenario according to which supermassive black hole pairs in gaseous environments tend to evolve toward tight binary (Dotti, Merloni & Montuori, 2015), it is rather plausible that dual AGN with separations of tens of parsecs at $z \geq 0$ are more common than previously thought. However, detections of unambiguous sub-kpc binaries at $z \geq 0$ are still elusive, even in the favourable case in which both supermassive black holes are accreting (Rodriguez et al., 2006; Bogdanović, Eracleous & Sigurdsson, 2009; Dotti et al., 2009; Eracleous et al., 2012). Planned X-ray observatories such as ATHENA+ will not have enough resolution to serve this purpose. However, the dense gaseous and dusty circumnuclear discs surrounding the binary would cause heavy absorption and scattering of the radiation emitted by active galactic nuclei to much longer wavelengths, possibly yielding emerging signals in the far infrared and radio wavelengths. With the current configuration ALMA may detect such binaries up to

$z \sim 0.2$, while JVLA and VLBI/eMERLIN, with their higher angular resolution of about 50 milliarcseconds should be able to probe sub-kpc binaries up to $z \sim 1$ or slightly higher, but would be biased to detect the brightest sources. At $z > 2$, gravitational wave observatories, such as the planned eLISA, will have the highest potential to provide indirect constraints on the population of supermassive black hole binaries.

5

Summary and future perspectives

The reciprocal roles of galaxies and supermassive black holes in their formation and mutual evolution has been the focus of extensive work during the last twenty years, yet a complete understanding of all the dominant processes is still missing and many questions remain open. The content of this Thesis represents an attempt to shed some light on a few relevant questions: how high-redshift galaxies have assembled? How supermassive black holes have formed? How supermassive black hole binaries evolve?

Although the answers to these questions in their full generality are not exhaustive yet, our results suggest possible solutions as well as new paths to be probed. Galaxies at $z \sim 2$ are often observed to have peculiar morphologies characterised by giant clumps with intense star formation, very different from those in the Local Universe. However, the targeted high-z objects are typically massive galaxies ($M_\star \sim 10^{11}$ M$_\odot$) with high star formation rates and gas fractions. They likely live in massive haloes continuously fed by large-scale accretion through cosmic filaments, and they might evolve into the most massive galaxies at the centre of present day galaxy clusters (or isolated giant ellipticals). Nonetheless, the evolutionary conditions of the underlying population of less massive galaxies might be different. This is shown in Chapter 2 through the results of the Argo simulation, which tracks the assembly of a common group down to $z = 3$, allowing us to study the evolution of a couple of dozens of galaxies with masses $M_\star \sim 10^{8-11}$ M$_\odot$ and characterised by a variety of morphologies and assembly histories. We find that a morphological diversity very similar to the Hubble Sequence, especially for late-type objects, is developed and in place at redshifts as high as $z \approx 3$, as indicated by several qualitative and quantitative indicators, such as the the stellar images, the surface density profiles, the B/T ratios and the distribution of angular momentum. While the lightest disc-like galaxies typically reside in peripheral haloes that never interact with the main group halo, the progressively more prominent bulges in the larger disc galaxies mostly form after the central starburst was triggered by a major merger. In some cases, tidal interactions due to fly-byes/minor mergers or the development of a bar can also lead to the growth of a bulge, but we do not find any relevant contribution from giant clumps inspiraling and merging at the galaxy centre. This suggests that the processes shaping the structure and morphology of low-to-moderate stellar mass galaxies at $z \gtrsim 3$ might proceed in a similar fashion to local counterparts.

The cores of most (massive) galaxies in the Universe host a supermassive black hole, whose formation process is still unknown. Many ideas have been proposed to address their formation process, mainly to explain the existence of very bright (and massive) AGNs at $z \sim 6-7$. Amongst them, the direct collapse scenario (see Section 1.2.4 and 3.1 for a more detailed description) may be a promising solution to the problem of very early and massive black holes. However, it somehow relies (at least in some variations) on the existence of exotic objects such as supermassive stars (i.e. radiation-dominated main sequence stars with masses $\sim 10^{5-6}$ $\mathrm{M}_{\odot}$) that evolve in quasi-stars, i.e. massive envelopes sustained against their own self-gravity by the accretion power of a central embryo black hole, the latter being able to grow very quickly to $\gtrsim 10^{4}$ $\mathrm{M}_{\odot}$. However, as we have shown in Chapter 3, the mass of the seed is conditional to the production of massive outflows which can be even faster than accretion in removing mass from the envelope. We have improved on previous models both on the description of the winds and on their relevance for quasi-stars. Specifically, we have shown that photon-tiring is not expected to occur in radiation-dominated winds driven by super-Eddington luminosities. This effect might appear only because of the lack of the enthalpy contribution in the governing equations. This might have important implications for quasi-stars: since the winds and the photosphere are bright, quasi-stars might be observable by future space-based telescope such as JWST. As a second step, we explore the role of rotation in quasi-stars within the simplifying assumptions of steady-state slow rotation. We apply a model of rotation in convective zones (originally developed for the Sun) to probe the typical angular velocity and momentum of the gas close to the Bondi radius of the black hole. We find that in most cases when the winds would not be a bottleneck for the black hole growth, the angular momentum is not enough to self-consistently allow the formation of an accretion disc outside the black hole horizon. This implies that the black hole would not provide enough feedback to reach equilibrium (i.e. to form a quasi-star) and therefore to halt the collapse of the surrounding envelope, which would mostly swallowed by the embryo black hole. Despite the limitations of our approach, we then speculate the existence of a bimodal population of seed black holes forming from direct collapse: when a quasi-star cannot form, a seed $\sim 10^{4-5}$ $\mathrm{M}_{\odot}$ might form in principle; on the other hand, when a quasi-star can form, the black hole power likely triggers massive outflows that prevent the black hole growth, leaving behind a $\sim 10^{2-3}$ $\mathrm{M}_{\odot}$ seed.

The assembly of galaxies proceeds through smooth accretion and mergers of smaller structures. When the latter involve progenitors already hosting supermassive black holes at their centres, the formation of supermassive black hole binaries appears as a natural prediction of the current cosmological paradigm. However, the details and the efficiency of the process subtly depend on the structure of the host galaxies and of the merger remnant, e.g. whether it is a minor or a major merger, whether it is gas-rich or not, and so on. In Chapter 4, we focus on gas-rich major mergers, first by means of idealised simulation of circumnuclear discs and then with self-consistent galaxy mergers. Specifically, we look at the effects that an inhomogeneous and clumpy interstellar medium may have on the dynamics of massive black hole pairs at separations of about 200 pc and below. We find that several effects may stochastically enlarge the range of orbital decay timescales allowed from ~ 1 up to ~ 100 Myr. The black holes can either receive angular momentum from dense spiral arms just internal to their orbit or scatter off gaseous clouds with masses similar to or larger then then the black hole mass. In the latter case, they might sometimes either get kicked toward the centre or get ejected outside the dense gas of the circumnuclear disc, mostly wandering within the surrounding stellar bulge, which in turn provides a weaker dynamical friction and therefore induces a longer decay timescale. We also estimate that such effect might be important for black holes of about 10^{6-7} $\mathrm{M}_\odot$ at $z \sim 2-5$, i.e. in the redshift and mass windows of the space-based gravitational wave observatory eLISA. Such a prediction seems to be confirmed by recent work on black hole pairs in high-z galaxies (Tamburello et al., 2016).

Future perspectives

While trying to answer some scientific questions, the investigations presented in this Thesis also open further questions.

The direct collapse scenario seems really intriguing to explain the formation of massive seeds that could represent fundamental bricks in the formation process of bright quasars at high-z. Nonetheless, it is clear that the details of such model still have to be worked out further. Cosmological simulations have been fairly successful in reproducing the large scale conditions for the direct collapse to take place, although the debate is still ongoing. Once a $\sim 10^6$ $\mathrm{M}_\odot$ cloud of about a parsec in size forms, it is really not clear what happens next. The common view is that a supermassive star would form, followed either by a quasi-stars or by the collapse of whole structure due to general relativistic radial instability. However, the results of Chapter 3, though they should be taken with care since they are based on a simple model, pose doubts about the stability and self-consistency of the quasi-star model. These issues are worth to be explored with numerical simulations. Despite that a fully self-consistent model of a quasi-star with radiation-hydrodynamics resolving accretion region is still beyond the capability of any simulation code, we are planning to perform some idealised simulations of spherical envelops with a central source of energy (implementing e.g. equation (3.26) as a boundary condition) to model the convective envelope of the quasi-star, where hydrodynamical effects are dominating over radiation effects since the transport of energy occurs through convection.

Once a black hole is formed, it mostly grows by accretion of gas from the surrounding galaxy, and also by mergers with other black holes. Therefore, modelling in details the host galaxy at $z > 5 - 6$ is of paramount importance to understand the ultimate feeding mechanism of growing black holes in the early Universe. We have already started a project to explore the properties of the interstellar medium of high-redshift galaxies, but without considering the feedback effect of a central black hole. Therefore, an important follow-up would be to include black hole physics in very high resolution cosmological simulations of $z \gtrsim 6$ galaxies to probe the growth of the central black holes, possibly exploring different accretion schemes. The same approach would naturally apply to study the dynamics of binary black holes in cosmological mergers of $z \sim 2 - 5$ gas-rich galaxies, to determine the role of the interstellar medium in influencing the dynamics of shrinking pairs and in oder to better constraint the typical coalescence timescale in the right conditions for future space-based observatories, such as eLISA.

A

Numerical methods: a brief review

In this Chapter, I will briefly describe the main features of the numerical techniques adopted throughout the Thesis. I will first give a general introduction of the equations that are solved by discretising the system in N bodies, and then I will describe the smoothed particle hydrodynamics (SPH) approach for hydrodynamics, mostly referring to the GASOLINE code, widely used for the calculations presented in this Thesis.

A.1 N-body/hydrodynamical simulations: basic concepts

The time evolution of astrophysical systems (e.g. galaxies) can be modelled with N-body numerical simulations. An N-body simulation computes the motion of particles that interact with one another through some type of physical forces. Particles usually represent discretised mass elements of the system under consideration.

Astrophysical systems can be generally divided in *collisional* (e.g. stellar clusters) and *collisionless systems* (e.g. stars in galaxies). In the following, we will focus on collisionless systems. Collisionless systems are characterised by the following relation between the *relaxation time* $t_{\rm relax}$ and the *crossing time* $t_{\rm cross}$:

$$t_{\rm relax} \equiv \frac{N}{8 \ln \Lambda} \frac{R}{\sigma} \gg t_{\rm cross} \equiv \frac{R}{\sigma}, \tag{A.1}$$

where N is the number of interacting stars with typical velocity σ determined by the smooth distribution of mass of the system, R is the scale length of the system, and $\ln \Lambda$ is the Coulomb logarithm, i.e. the logarithm of the ratio between the largest and the smallest scales of interaction between stars (typically, $\ln \Lambda \sim \ln N$). $t_{\rm cross}$ gives the timescale for a star to cross the system under the influence of the global gravitational potential. Instead, $t_{\rm relax}$ is the timescale on which kinetic energy of a star changes significantly because of two-body encounters with other stars. Therefore, the condition of equation (A.1) states that the dynamics of a collisionless system is specified by the overall smooth mass distribution and gravitational potential of the system (i.e. two-body relaxation is negligible). A typical galaxy is made of $N \sim 10^{11}$ stars, which means $N/(8\ln(N)) \sim 10^8 - 10^9$, and has an age $\sim 100\ t_{\rm cross}$, therefore its stellar component is a collisionless system.

The time evolution of a collisionless system is described by the collisionless Boltzmann equation (CBE; see e.g. Binney & Tremaine 2008):

$$\frac{\partial f}{\partial t} + (\mathbf{v} \cdot \nabla) f - \nabla\Phi \cdot \frac{\partial f}{\partial \mathbf{v}} = 0, \tag{A.2}$$

self-consistently coupled with the Poisson equation:

$$\nabla^2\Phi(\mathbf{x}, t) = 4\pi G\ \rho(\mathbf{x}, t) = 4\pi G \int f(\mathbf{x}, \mathbf{v}, t)\ \mathrm{d}^3\mathbf{v}, \tag{A.3}$$

where $\rho = \int f\ \mathrm{d}^3\mathbf{v}$ is the density field of the system, $f(\mathbf{x}, \mathbf{v}, t)$ is the distribution function in the phase-space $\mathcal{P}(\mathbf{x}, \mathbf{v})$, and $\Phi(\mathbf{x}, t)$ is the gravitational potential. The distribution function gives the mass of stars $\mathrm{d}M = f(\mathbf{x}, \mathbf{v}, t)\ \mathrm{d}^3\mathbf{x}\ \mathrm{d}^3\mathbf{v}$ inside the phase-space volume $\mathrm{d}^3\mathbf{x}\ \mathrm{d}^3\mathbf{v}$. Considering a discrete system of N stars, for the i-th star equation (A.2) reads:

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} = \mathbf{v}_i, \tag{A.4}$$

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\nabla\Phi(\mathbf{x}_i, t). \tag{A.5}$$

Numerically, a collisionless system is treated as a set of N particles with mass m_i that evolves in time according to equations (A.4) and (A.5) under the gravitational potential calculated following equation (A.3).

The CBE describes a continuous system without two-body interactions, whereas a couple of particles in a discrete system can in principle approach each other up to the point when the Newtonian force between them becomes large and diverges. Such a divergence is an artefact of the discrete sampling of the density distribution. Therefore, the acceleration in equation (A.5) needs to be modified in order to avoid two-body interactions. The gravitational potential that determines the acceleration of the i-th particle at $\mathbf{x}_i$ is modified by means of the *softening kernel* function $\mathcal{S}(\mathbf{x}, \epsilon)$:

$$\Phi(\mathbf{x}_i, t)\ \rightarrow\ \widetilde{\Phi}_i \equiv \widetilde{\Phi}(\mathbf{x}_i, t; \epsilon) \equiv \int \Phi(\mathbf{y}, t)\ \mathcal{S}(\mathbf{x}_i - \mathbf{y}, \epsilon)\ \mathrm{d}\mathbf{y}. \tag{A.6}$$

$\mathcal{S}(\mathbf{x}, \epsilon)$ depends on the *gravitational softening length* ϵ, the scale length below which the gravitational interaction is reduced to avoid the effects of two-body encounters between very close particles. This modification is equivalent to reducing the value of the Coulomb logarithm for a given N, increasing the proportionality factor between $t_{\rm relax}$ and $t_{\rm cross}$.

Another important component to model is gas. Gas is treated as a perfect fluid coupled with the collisionless components through gravity. It is described by the following equations:

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) &= 0, && \text{(A.7)}\\
\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} + \frac{\nabla P}{\rho} + \nabla \Phi &= 0, && \text{(A.8)}\\
\frac{\partial u}{\partial t} + (\mathbf{v} \cdot \nabla) u + \frac{P}{\rho}\ \nabla \cdot \mathbf{v} &= 0, && \text{(A.9)}
\end{aligned}
$$

where $\rho(t, \mathbf{x})$ is the gas density, $\mathbf{v}(t, \mathbf{x})$ is the gas velocity, $\Phi(t, \mathbf{x})$ is the gravitational potential, $u(t, \mathbf{x})$ is the internal energy per unit mass and P is the gas pressure. Equations (A.7), (A.8) and (A.9) describe the conservation of mass, momentum, and energy, respectively. These equations are closed by the equation of state $P = P(\rho, u) = (\gamma - 1)\ \rho\ u$, where γ is the adiabatic index. Several numerical schemes have been developed to solve these equations; they are mostly divided in two main groups: particle-based and grid-based schemes. The first approach uses point-like particles that represent mass elements of the fluid in a *Lagrangian* fashion, i.e. the motion of a single fluid element is followed both in space and time. On the other hand, grid-based methods follow the time evolution of the fluid on a discretised domain in space. This approach is *Eulerian* since it tracks the evolution with time of the fluid properties in each point in space.

A.2 Smoothed particle hydrodynamics: GASOLINE

GASOLINE is an N-body/hydrodynamical code that includes SPH (Gingold & Monaghan, 1977; Lucy, 1977) and K-D tree algorithm (Bentley, 1979) in order to solve the hydrodynamics and to calculate the gravitational forces, respectively (Wadsley, Stadel & Quinn, 2004).

The basic assumption behind SPH is that each local property of the fluid f in the position $\mathbf{x}$ can be computed averaging over the nearby region with a weight function $\mathcal{W}$ called *smoothing kernel*, that is:

$$
\langle f \rangle(\mathbf{x}) = \int f(\mathbf{x}')\ \mathcal{W}(\mathbf{x} - \mathbf{x}'; h)\ \mathrm{d}^3\mathbf{x}', \qquad \text{(A.10)}
$$

where h is the *smoothing length*, i.e. the scale length beyond which the kernel becomes negligible. Since SPH is a particle representation of the fluid, the integration in equation (A.10) becomes a (volume-weighted) summation over particles inside h:

$$
f_{\mathrm{i}} = \sum_{j=1}^{N_i} f_j \frac{m_j}{\rho_j}\ \mathcal{W}_{ij}(x_{ij}, h_i, h_j), \qquad \text{(A.11)}
$$

where the summation extends to the N_i neighbours of the i-th particle, whose distances $x_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ are less then $2h_i$, i.e. two times the smoothing length. The quantities with the j subscript are associated with the j-th particle, while $\mathcal{W}_{ij}(\mathbf{x}_{ij}, h_i, h_j)$ stands for the symmetric kernel:

$$W_{ij}(\mathbf{x}_{ij}, h_i, h_j) = \frac{1}{2} w(x_{ij}/h_i) + \frac{1}{2} w(x_{ij}/h_j), \tag{A.12}$$

that is necessary for energy and momentum conservation (Hernquist & Katz, 1989; Wadsley, Stadel & Quinn, 2004). The function $w(x)$ is a normalised spherical kernel with compact support (Monaghan, 1992):

$$w(q) = \frac{1}{\pi h^3} \begin{cases} 1 - (3/2)\, q^2 + (3/4)\, q^3, & 0 \leq q \leq 1, \\ (2-q)^3/4, & 1 < q \leq 2, \\ 0, & q > 2. \end{cases} \tag{A.13}$$

Each particle represents a fluid element with mass m_i. Thus, the density of each SPH particle is:

$$\rho_i = \sum_{j=1}^{N_i} m_j\ \mathcal{W}_{ij}(x_{ij}, h_i, h_j). \tag{A.14}$$

This equation represents the conservation of mass, equation (A.7). Variable smoothing lengths are commonly used in order to maintain N_i at a nearly constant value $N_{\rm sph}$. Following this idea, the SPH implementation of the momentum and energy equations for particle i, equations. (A.8) and (A.9), is:

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_{j=1}^{N_{\rm sph}} m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \nabla_i \mathcal{W}_{ij} + \mathbf{f}_{{\rm grav},i}, \tag{A.15}$$

and

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_{j=1}^{N_{\rm sph}} m_j \left(\frac{P_i}{\rho_i^2} + \frac{1}{2}\Pi_{ij} \right) \mathbf{v}_{ij} \cdot \nabla_i \mathcal{W}_{ij}, \tag{A.16}$$

where $\mathbf{v}_{ij} \equiv \mathbf{v}_i - \mathbf{v}_j$, $\mathcal{W}_{ij} \equiv \mathcal{W}_{ij}(x_{ij}, h_i, h_j)$ for brevity, $\mathbf{f}_{{\rm grav},i}$ is the gravitational force on particle i, and Π_{ij} is a term of artificial viscosity. The artificial viscosity, which is necessary to capture shocks, is implemented in GASOLINE as:

$$\Pi_{ij} = \begin{cases} \dfrac{-\alpha(c_i + c_j)\mu_{ij}/2 + \beta \mu_{ij}^2}{(\rho_i + \rho_j)/2} & \mathbf{x}_{ij} \cdot \mathbf{v}_{ij} < 0, \\ 0 & \mathbf{x}_{ij} \cdot \mathbf{v}_{ij} \geq 0, \end{cases} \tag{A.17}$$

where c_j is the sound speed of the j-th particle, $\alpha = 1$ and $\beta = 2$ are coefficient representing shear viscosity and a Von Neumann-Richtmyer-like viscosity (for high Mach numbers) respectively, and finally:

$$\mu_{ij} = \frac{(h_i + h_j)\mathbf{x}_{ij} \cdot \mathbf{v}_{ij}/2}{x_{ij}^2 + 0.01(h_i + h_j)^2}. \tag{A.18}$$

In addition, the artificial viscosity term is multiplied by the Balsara switch, $b = |\nabla \cdot \mathbf{v}|/(|\nabla \cdot \mathbf{v}| + |\nabla \times \mathbf{v}|)$, that limits the viscosity in non-shocking, shearing environments (Balsara, 1995).

The standard SPH implementation in GASOLINE has been further improved in order to alleviate some difficulties that traditional SPH has in capturing instabilities at the boundaries between regions with steep density gradients (Agertz et al., 2007). Being an exact Lagrangian method, SPH is not able to explicitly capture turbulent mixing between fluid elements below the coarse-grained volume represented by each particle, since no explicit advection is present. While artificial viscosity provides dissipation for the kinetic energy, Wadsley, Veeravalli & Couchman (2008) have introduced an additional explicit diffusive term in the energy equation (A.16) to model the unresolved heat flux due to turbulent dissipation:

$$\left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{dif}} = -\sum_{j=1}^{N_{\mathrm{sph}}} m_j Q_{ij}(u_i - u_j)\frac{\mathbf{x}_i - \mathbf{x}_j}{x_{ij}^2}\nabla_i \mathcal{W}_{ij}. \tag{A.19}$$

The diffusive coefficient Q_{ij} is given by:

$$Q_{ij} = C\frac{|\mathbf{v}_i - \mathbf{v}_j|\;(h_i + h_j)}{\rho_i + \rho_j}, \tag{A.20}$$

where the dimensionless coefficient $C \approx 0.05$ is an adjustable parameter. This model has been shown to improve in the comparison between SPH and grid codes and produces better results in problems where the role of turbulent diffusive processes is important, such as the entropy distribution in the cores of galaxy clusters (Wadsley, Veeravalli & Couchman, 2008). The model has been further improved by Shen, Wadsley & Stinson (2010), who also apply it to the unresolved turbulent mixing of metals.

Both the gaseous and the collisionless components (e.g dark matter) interact gravitationally. GASOLINE computes the gravitational interactions adopting a strategy similar to that proposed by Barnes & Hut (1986), but using a K-D tree instead of an oct-tree. Starting from the main domain of the simulation, the gravity K-D tree is built by recursively bisecting the longest axis of each cell. At each level, the physical dimensions of the cells are squeezed to just contain the particles. The process is halted when $n_{\mathrm{bucket}} \approx 8-16$ particles are contained within leaf cells. Then, the centre of mass and the multipole moments of the mass distribution up to the hexadecapole (i.e. 4^{th} order) are computed for each cell, starting from the leaf up to the root cell (i.e. the main domain).

Gravity is computed by "walking" the tree from the root and assigning to each cell the opening radius r_{open}:

$$r_{\mathrm{open}} = \frac{2L}{\sqrt{3}\theta}, \tag{A.21}$$

where L is the largest distance between the centre of mass of the cell and a particle within the cell, and $\theta < 1$ is an accuracy parameter. When the sphere centred on the centre of mass of a cell C_j with radius $r_{\mathrm{open},j}$ intersects the volume of the leaf cell B_i, the cell C_j is opened in the cell's children and each of them are analysed in the same way. This operation is recursively repeated until: (i) the sphere centred on C_j does not intersect the volume of B_i, and therefore the contribution of C_j to the gravitational force acting on the particles in B_i is computed from the multipole expansion; or (ii) the tree walking reaches another leaf cell B_j, and the gravitational force contribution of the particles within B_j is added directly to the gravitational force acting on each particle inside B_i. In case (ii), the gravitational forces are smoothed as described by equation (A.6), where the gravitational kernel $\mathcal{S}$ has the same mathematical form as the SPH kernel in equation (A.13), using the smoothing length $h = \epsilon_i + \epsilon_j$, where ϵ_i is the gravitational softening of the i-th particle.

B

Additional papers

In the following, we list and attach some additional papers that we contributed to during the development of this Thesis. In chronological order, the papers are:

1. Mayer, L., **Fiacconi, D.**, et al., *Direct formation of supermassive black holes in metal-enriched gas at the heart of high-redshift galaxy mergers*, 2015, ApJ, 810, 51
2. Fanali, R., Dotti, M., **Fiacconi, D.**, & Haardt, F., *Bar-formation as driver of gas inflows in isolated disc galaxies*, 2015, MNRAS, 454, 3641
3. Lupi, A., Haardt, F., Dotti, M., **Fiacconi, D.**, Mayer, L., & Madau, P., *Growing massive black holes through super-critical accretion of stellar-mass seeds*, 2016, MNRAS, 456, 2993
4. **Fiacconi, D.**, Madau, P., Potter, D., & Stadel, J., *Cold dark matter substructures in elliptical galaxy halos*, accepted for publication in ApJ, arXiv:1602.03526
5. Khan, F., **Fiacconi, D.**, Mayer, L., Berczik, P., & Just, A., *Swift coalescence of supermassive black holes in cosmological mergers of massive galaxies*, submitted to ApJ Letters, arXiv:1604.00015

Preprint typeset using LaTeX style emulateapj v. 12/16/11

DIRECT FORMATION OF SUPERMASSIVE BLACK HOLES IN METAL-ENRICHED GAS AT THE HEART OF HIGH-REDSHIFT GALAXY MERGERS

Lucio Mayer[1], Davide Fiacconi[1], Silvia Bonoli[2], Thomas Quinn[3], Rok Roskar[4], Sijing Shen[5], James Wadsley[6]

ABSTRACT

We present novel 3D multi-scale smoothed particle hydrodynamics (SPH) simulations of gas-rich galaxy mergers between the most massive galaxies at $z \sim 8-10$, designed to scrutinize the direct collapse formation scenario for massive black hole seeds proposed in Mayer et al. The simulations achieve a resolution of 0.1 pc, and include both metallicity-dependent optically-thin cooling and a model for thermal balance at high optical depth. We consider different formulations of the SPH hydrodynamical equations, including thermal and metal diffusion. When the two merging galaxy cores collide, gas infall produces a compact, optically thick nuclear disk with densities exceeding 10^{-10} g cm^3. The disk rapidly accretes higher angular momentum gas from its surroundings reaching ~ 5 pc and a mass of $\gtrsim 10^9\ M_\odot$ in only a few 10^4 yr. Outside $\gtrsim 2$ pc it fragments into massive clumps. Instead, supersonic turbulence prevents fragmentation in the inner parsec region, which remains warm ($\sim 3000-6000$ K) and develops strong non-axisymmetric modes that cause prominent radial gas inflows ($> 10^4\ M_\odot$ yr^{-1}), forming an ultra-dense massive disky core. Angular momentum transport by non-axisymmetric modes should continue below our spatial resolution limit, quickly turning the disky core into a supermassive protostar which can collapse directly into a massive black hole of mass $10^8-10^9\ M_\odot$ via the relativistic radial instability. Such a "cold direct collapse" explains naturally the early emergence of high-z QSOs. Its telltale signature would be a burst of gravitational waves in the frequency range $10^{-4}-10^{-1}$ Hz, possibly detectable by the planned *eLISA* interferometer.

Subject headings: Black hole physics – Galaxies: nuclei – Hydrodynamics – Methods: numerical

1. INTRODUCTION

The origin of supermassive black holes (SMBHs) is still a puzzle. At least three different classes of models have been proposed: light seeds from Pop III stars, massive seeds from direct gas collapse and light to intermediate mass seeds formed inside star clusters via runaway mergers of stars and stellar remnants. Light seeds from Pop III stars originate at $z > 12-15$ and are the natural consequence of primordial star formation (Madau & Rees 2001; Volonteri et al. 2003; Tanaka & Haiman 2009). However, they may face difficulties in growing at a fast enough rate, at least comparable to the Eddington rate, to produce black holes exceeding $10^9\ M_\odot$ in less than half a billion years (Johnson & Bromm 2007; Pelupessy et al. 2007; Wise et al. 2008; Alvarez et al. 2009; Milosavljević et al. 2009), as required by the existence of high-z quasars (QSOs; Fan et al. 2001; Mortlock et al. 2011; Treister et al. 2013). Pop III seeds indeed tend to grow at rates well below Eddington as they are surrounded by low density ionized gas and are hosted in small halos that cannot sustain fast inflows of fresh gas from the cosmic web due to their shallow potential wells. Seeds from dense star clusters form when the cluster is subject to rapid mass segregation and consequent runaway mergers that lead to the formation of a central very massive star $\gtrsim 1000\ M_\odot$ (VMS; Gürkan et al. 2004; Freitag et al. 2006). The VMS is then expected to end its life leaving behind a BH seed with $\sim 100-1000\ M_\odot$ (see e.g. Portegies Zwart et al. 2004; Belkus et al. 2007). However, intense stellar wind may strongly suppress the growth of the VMS (e.g. Vink 2008; Glebbeek et al. 2009), unless it evolves in low metallicity environments such as high-z nuclear star cluster (Devecchi & Volonteri 2009; Devecchi et al. 2012). The resulting BH seed might be more massive if the entire star cluster is pushed towards rapid collapse by mass loading from a large scale gas inflow (Davies et al. 2011; Lupi et al. 2014). Finally, direct collapse models rely on gravitational instabilities, such as spiral instabilities and bars-in-bars, to transport gas all the way from galactic disk scales to sub-pc scales (Begelman et al. 2006; Lodato & Natarajan 2006; Begelman & Shlosman 2009). As a result, a Jeans unstable cloud forms and it may later evolve into a supermassive star (SMS), which in turn produces a quasi-star (QS) nurturing a massive BH seed ($M \sim 10^4-10^5\ M_\odot$) at its center (Begelman et al. 2008; Volonteri & Begelman 2010; Dotan et al. 2011). Alternatively, if the cloud or supermassive star is massive enough ($M > 10^5-10^6\ M_\odot$), and has negligible rotational support, it can collapse via the general relativistic radial instability all the way into a black hole encompassing more than half of its original mass (Hoyle et al. 1963; Baumgarte & Shapiro 1999; Shibata & Shapiro 2002; Saijo & Hawke 2009; Montero et al. 2012). However, conventional direct collapse scenarios posit suppression of radiative cooling in order to avoid fragmentation and form a single supermassive object (Lodato & Natarajan 2006; Dijkstra et al. 2006). This in turn requires primordial

---

[1] Center for Theoretical Astrophysics and Cosmology, Institute for Computational Science, & Physik Institut, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

[2] Centro de Estudios de Fisica del Cosmos de Aragon, Plaza San Juan 1, Planta-2, 44001, Teruel, Spain

[3] Astronomy Department, University of Washington, Box 351580, Seattle, WA, 98195-1580, USA

[4] Scientific IT Services, ETH Zürich, Weinbergstrasse 11, CH-8092 Zürich, Switzerland

[5] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

[6] Department of Physics and Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada

gas composition and efficient dissociation of one of the most efficient primordial coolants, namely molecular hydrogen. These requirements can be fulfilled under special environmental conditions such as strong external Lyman-Werner ionizing flux (see Ferrara & Loeb 2013; Dijkstra et al. 2014). The conditions arising in direct collapse models may also restart the growth of a dormant Pop III seed that has been previously accreting inefficiently (Begelman 2012; Madau et al. 2014). Interestingly, cosmological simulations have shown convincingly that cold gas accretion flows originating from the cosmic web can sustain a mass accretion rate high enough to produce the high-z QSOs on the required short timescale in rare peaks at $z > 6$ (halos with virial mass $M_{\rm vir} \sim 10^{12} - 10^{13}$ $M_{\odot}$), provided that BH seeds in such halos start with a mass $> 10^5$ $M_{\odot}$ and that they accrete constantly near the Eddington rate (Di Matteo et al. 2012).

An alternative scenario for direct collapse has been proposed by Mayer et al. (2010, hereafter MA10), and later tested against observational constraints of $z > 6$ QSOs using a semi-analytical model attached to the *Millennium Simulation* by Bonoli et al. (2014). In such a scenario, massive BH seeds originate from the collapse of a supermassive cloud assembled at the end of $z \sim 4 - 10$ gas-rich galaxy mergers by multi-scale gas inflows driven by gravoturbulence and gravitational torques. The model does not require primordial gas composition, rather it is designed to be at work in very massive galaxies that have already been enriched to solar metallicity at $z < 10$. MA10 measure gas inflow rates $\dot{M} \sim 10^3 - 10^4$ $M_{\odot}$ yr^{-1} at parsec scales and below. The merger-driven model postulates that the formed seeds grow fast but are rare, since they require gas-rich, major mergers between highly biased host halos corresponding to $4 - 5\sigma$ fluctuation (Bonoli et al. 2014). On the small scale side, the model predicts inflows that well exceed those necessary to build a SMS (see Bonoli et al. 2014, and references therein), while other direct collapse models based on instabilities in protogalaxies at $z > 10$ fall short of the required inflow rates (e.g. Regan & Haehnelt 2009a,b; Latif et al. 2012).

The model, however, has been questioned since it is based on simulations with idealized thermodynamical conditions. Indeed, the simulations of MA10 employed an effective equation of state (EoS) with a variable polytropic index across a wide range of gas densities, valid for metal-enriched gas. It has been pointed out that the adopted EoS may artificially promote the formation of a central supermassive gas cloud by suppressing gas fragmentation and star formation. Instead, the expected high cooling rates should lead to widespread fragmentation, disrupting the gas inflow and leading to a BH of only ~ 100 $M_{\odot}$ in the most optimistic scenario (Ferrara et al. 2013). The latter arguments, however, are based on simple one-dimensional disk models evolved in the isochoric limit, which forces the density to remain constant at a given radius, neglecting the effect of shocks. These models also neglect self-gravity. Therefore they lack the self-consistent coupling between thermodynamics and hydrodynamics. It is conceivable that self-gravity is a crucial ingredient in the MA10 model since the inflow is generated by multi-scale gravitational torques and it is ultimately mediated by a massive, self-gravitating, marginally unstable nuclear gas disk at scales $\lesssim 100$ pc.

Here we present 3D hydrodynamic simulations with initial conditions identical to MA10 but with a much richer inventory of physical processes. Most notably, we include radiative cooling in both the optically-thin and optically-thick regimes in place of a prescribed EoS. We even explore the effect of different formulations of SPH, including a novel implementation of the SPH hydro force along with thermal diffusion which successfully obviate to the notorious difficulties of capturing fluid instabilities and mixing (Keller et al. 2014 – see also Ritchie & Thomas 2001 and Hopkins 2013).

As we show in this paper, the basic picture proposed in MA10 is confirmed and further strengthened. Not only the inflow occurs and forms a supermassive cloud even under these more realistic thermodynamical conditions, but it is enhanced by cooling, producing a much more compact and denser cloud on a shorter timescale. As a result, we are left with the intriguing possibility that the collapse proceeds all the way to direct SMBH formation via the relativistic radial instability. Despite the metal-line cooling, fragmentation is limited due to shock heating and gravoturbulence. We conclude that, contrary to previous claims, our scenario for direct collapse is tenable. It offers an explanation for the rapid emergence of high-z QSOs without requiring any special conditions in the interstellar medium (ISM), such as primordial gas composition and dissociation of molecular hydrogen.

The paper is organized as follows: in Section 2 we describe the setup of the simulations and the implementation of the physical processes included. We present our findings in Section 3, while we discuss possible evolution scenarios of our final configuration towards a massive BH seed in Section 4. In Section 5 we highlight the limitations that could affect our results and we present our conclusions.

2. NUMERICAL SIMULATIONS

2.1. *Initial conditions and particle splitting*

The main purpose of this work is to study the gas dynamics of a gas-rich major galaxy merger from kiloparsec scales down to ~ 0.1 pc. To this aim, we start by running a large scale galaxy merger simulation and then we apply particle splitting in the later phases of the evolution to reach sub-pc resolution, following MA10. Since we want to compare with our previous work, we adopt the same initial conditions of the reference large-scale merger simulation used in MA10, namely a prograde coplanar 1:1 merger. The merger simulation was initially run with a mass resolution of 2×10^4 $M_{\odot}$ for the gas particles and a gravitational softening of 100 pc for all galaxy components. The simulation includes radiative cooling, star formation and supernova feedback (see Kazantzidis et al. 2005; Mayer et al. 2007; MA10 for further details). The galaxies approach on a typical parabolic orbit inferred from cosmological simulations (Khochfar & Burkert 2006). The significant computational burden demanded by these simulations favors the choice of a coplanar prograde merger, which leads to full coalescence of the two galaxy cores on a timescale somewhat shorter than in the other configurations (Kazantzidis et al. 2005; Capelo et al. 2014). Nonetheless, the mass of gas that concentrates at scales $\gtrsim 100$ pc, which is a key feature in our scenario, has been shown to be rather insensitive to the merger geometry for a fixed mass ratio between the two galaxies (Kazantzidis et al. 2005).

The two galaxy models represent identical multi-component systems with virial mass $M_{\rm vir} = 10^{12}\ M_{\odot}$. They are constructed following the method by Hernquist (1993) and choosing the structural parameters in agreement with the scaling laws predicted by the Λ-CDM model (Mo et al. 1998; Kazantzidis et al. 2005). The models include a dark matter halo that follows the Navarro-Frank-White (NFW; Navarro et al. 1996) profile with concentration $c = 12$, an exponential stellar and gaseous disk and a stellar bulge. The stellar disk has mass $M_{\rm disk} = 6 \times 10^{10}\ M_{\odot}$, scale radius $R_{\rm disk} = 3.5$ kpc and scale height $z_{\rm disk} = 350$ pc. The gaseous disk has the same scale lengths of the stellar disk. The mass is about $0.1 M_{\rm disk}$ at the time of the last pericenter passage, when star formation has already consumed part of the gas. The bulge is a Hernquist (1990) model with mass $M_{\rm bulge} = M_{\rm disk}/5$ and scale length $r_{\rm bulge} = 0.7$ kpc. We refer to MA10 for more details on the numerical parameters of the large scale merger simulation (see also Kazantzidis et al. 2005; Mayer et al. 2007).

We note that the adopted virial mass is comparable to that of a Milky Way-sized galaxy at $z = 0$, and it corresponds to $\sim 5\sigma$ fluctuations at $z = 8 - 9$ in the *WMAP9* cosmology. This is consistent with the very low abundance and high clustering amplitudes of high-z QSOs (Volonteri & Rees 2006; Bonoli et al. 2014). Our galaxy models are similar to present-day, massive, disk-dominated galaxy. This is supported by recent cosmological simulations showing that typical disk dominated galaxies at $z > 3$ have morphologies already akin to their counterparts at low z, with a disk and a sizable bulge component (Fiacconi et al. 2015). However, gas fractions are often higher at high-z, in the range $0.3 - 0.6$, as suggested by both observations and simulations (Tacconi et al. 2010; Moody et al. 2014; Fiacconi et al. 2015), and the most massive among high-z galaxies might have hosted clumpy, turbulent disks (Tacconi et al. 2010; Ceverino et al. 2012; Förster Schreiber et al. 2014; Moody et al. 2014) rather than smooth exponential disks as in our initial conditions. Therefore our initial conditions follow a conservative approach since a more gas-rich, compact, and gravoturbulent disk will promote loss of angular momentum and central inflow even before the merger occurs (e.g. Bournaud et al. 2012).

As in MA10, we split individual gas particles within a spherical volume of radius 30 kpc into 8 child particles, when galaxy cores are 6 kpc apart at their last apocenter. We reduce correspondingly their gravitational softening, achieving a resolution of $\sim 2600\ M_{\odot}$ and 0.1 pc. Preexisting star particles and dark matter particles are not split, and their softening is left unvaried. They essentially provide a smooth background potential to avoid spurious two-body heating against the much lighter gas particles. The momentum conserving splitting procedure, which improves over that used in MA10, is described in Roškar et al. (2015). The choices of when to split the gas particles and the size of splitting volume are identical to that in MA10 and are discussed in the Supplementary Information of the latter paper. They are motivated by inducing minimal numerical fluctuations by introducing a refined region large enough to avoid any contamination of low-res particles for the entire duration of the simulations. After the two cores reached a separation < 100 pc, most of the gaseous mass is collected in the inner 100 pc volume and it is traced with as many as 1.5 million gas particles. By performing numerical tests we have verified that, owing to the fact that gas dominates the mass and dynamics of the nuclear region, the large softening adopted for the dark matter particles does not affect significantly the density profile of the inner dark halo that surrounds the nuclear disk. These and other numerical tests showing the robustness of the technique were presented in Mayer et al. (2007).

2.2. *Features of the simulation suite*

We have run four different versions of the second part of the simulation (i.e. after particle splitting) using GASOLINE2, a new version of the GASOLINE code (Wadsley et al. 2004; Keller et al. 2014). Individual runs differ in the sub-grid model for radiative cooling as well as for the specific implementation of the SPH equations of hydrodynamics. We describe below the different features of the runs and we summarize them in Table 1. In the following, we refer to each run with its own label according to Table 1. In all the runs we adopt the metal-dependent, optically-thin cooling introduced in Shen et al. (2010). It considers tabulated cooling rates in ionization equilibrium, while for H and He we compute directly the rates without assuming equilibrium (Wadsley et al. 2004). As in MA10, we assume solar metallicity gas, consistent with observational constraints on the metallicity of the hosts of high-z QSOs (Walter et al. 2004). In the following, we call this scheme “metal cooling”. In one of the runs (RUN3), the code switches to an equilibrium temperature-density relation above $\rho_{\rm g} = 0.1$ H cm^{-3}. This relation is calibrated on 2D radiative transfer calculations based on an improved version of the Spaans & Silk (2000) model, as described in detail in Roškar et al. (2015). In the following, we call this additional feature “thermal balance model”. The “thermal balance model” includes: (1) cooling by molecular lines and by collisions between dust and gas with metallicity-dependent opacity effects due to absorption and scattering of photons by dust; (2) IR dust radiation; (3) photoelectric effect on dust; (4) atomic and molecular line trapping in an ISM irradiated by stellar light; and finally (5) heating by cosmic rays. The thermal balance model thus accounts for self-shielding effects in the dense ISM. We note that one process that we do not take into account is the optically-thin cooling by H_2 molecules. This would be very important at low densities in low-metallicity, primordial halos, but in the metal-enriched system under study the cooling is dominated by fine structure metal lines, which are included in our “metal cooling” module. Indeed while typical cooling rates are of order $10^{-27} - 10^{-26}$ erg cm^3 s^{-1} due to fine structure metal lines at temperatures of a few thousand K, the cooling rates for H_2 in the high density limit ($> 10^8$ H cm^{-3}), which approaches local thermodynamic equilibrium, are only of order 10^{-28} erg cm^3 s^{-1}, hence significantly lower (Hollenbach & McKee 1979; Galli & Palla 1998).

The “thermal balance model” assumes the presence of a uniform UV photon radiation field produced by a starburst in gas of a specified metallicity. Before the final

TABLE 1
List of performed simulations with their characteristics

Label	SPH force	cooling type	thermal diffusion	metal diffusion
RUN1	standard	metal cooling	no	no
RUN2	standard	metal cooling	yes	yes
RUN3	standard	metal cooling + thermal balance model	yes	yes
RUN4	GDSPH	metal cooling	yes	yes

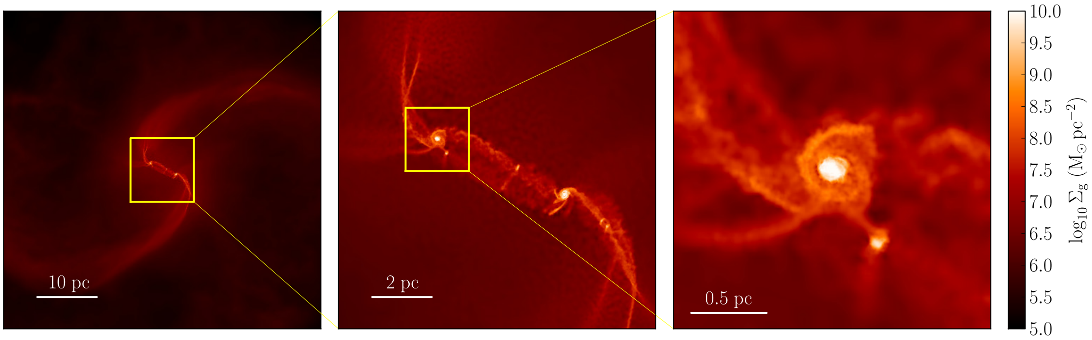

Fig. 1.— Projected gas density of RUN4 taken just prior to the final merger of the two cores, at time $t = t_0 - 3.2$ kyr, where t_0 is the time of coalescence of the approaching cores (see the text for details). The zoom-in view to the inner few parsecs shows one of the two dense cores and a few clumps produced by fragmentation of the infalling gas.

merger, the large scale simulation show that a starburst with a strength of $\sim 100\ M_\odot\ \mathrm{yr}^{-1}$ takes place in the inner kpc (see MA10). Therefore, we assume such a star formation rate as input to determine the stellar UV flux required by the thermal balance model. While the intensity of the starburst is actually a free parameter, the assumed value is on the low side. A starburst with ten times higher strength, which is likely at high-z, would produce a higher UV flux intensity, which in turn would heat up the dust and enhance photoelectric heating of the gas; this would produce a higher gas temperature floor and moderate net radiative losses, strengthening further the main findings of this paper concerning stability to fragmentation (see Section 3).

A pressure floor ensures that the Jeans mass is resolved locally by $\gtrsim 1$ resolution element, in order to avoid spurious fragmentation. Following Agertz et al. (2009), the minimum pressure is set to $P_{\mathrm{min}} = \alpha G \max(h, \epsilon)^2 \rho^2$, where $\alpha = 3.0$ is a fudge factor, G is the gravitational constant, ρ is the density of the particle, and $\max(h, \epsilon)$ is the local resolution element, i.e. the maximum between the smoothing length h and the softening length ϵ of the gas particle. All the simulations (including the the large-scale merger simulation before splitting is applied) adopt star formation and blast-wave supernova feedback following Stinson et al. (2006). In particular, we allow stars to form from gas following a Schmidt law in regions above the density threshold $n_{\mathrm{th}} = 10^4$ H cm^{-3} and below the temperature threshold $T_{\mathrm{th}} = 500$ K.

We also compare runs with different implementations of SPH that should capture hydrodynamic instabilities and multi-phase fluid mixing with increasing degree of realism. In particular: (i) RUN1 employs the original SPH implementation of GASOLINE (dubbed in the following "standard" SPH; Wadsley et al. 2004); (ii) RUN2 and RUN3 employ the "standard" SPH with the thermal and metal diffusion terms in the momentum and energy equations based on the sub-grid turbulence prescription described by Shen et al. (2010); and (iii) RUN4 employs thermal and metal diffusion as well as the new implementation of the hydrodynamical force equation described by Keller et al. (2014). The new approach is based on the geometric density average (GDSPH) in the SPH force expression $(P_i + P_j)/(\rho_i \rho_j)$ in place of the usual $P_i/\rho_i^2 + P_j/\rho_j^2$, where P_i and ρ_i are particle pressures and densities respectively. This modification leads to smoother gradients and removes artificial surface tension (Keller et al. 2014). Detailed tests of the GDSPH implementation combined with diffusion as in RUN4 will be presented in Wadsley et al. (in preparation), showing that it can successfully capture Kelvin-Helmoltz and Rayleigh-Taylor instabilities. Note that other new SPH implementations, while they often track entropy rather than energy as we do, also use a geometric density mean for the forces (e.g. Hopkins 2013; Saitoh & Makino 2013).

3. RESULTS

3.1. *Global evolution of the nuclear region*

We begin our analysis when the two galaxy cores are only ~ 10 pc apart and $\lesssim 10$ kyr away from final coalescence, as shown by the gas density projection in Figure 1. As the two galaxy cores approach their last encounter, efficient cooling leads to some fragmentation within the colliding gas filaments surrounding them. However, most of the gas flows inward, and the two dense cores sink owing to dynamical friction against the diffuse background of gas and stars (the contribution of dark matter is negligible inside a radius of 100 pc). Hereafter we define the reference merging time of the cores, t_0, as the time at which two separate density peaks cannot be identified any longer. Our analysis will focus on $t \geq t_0$.

The mean gas flow is radial, with velocities exceeding

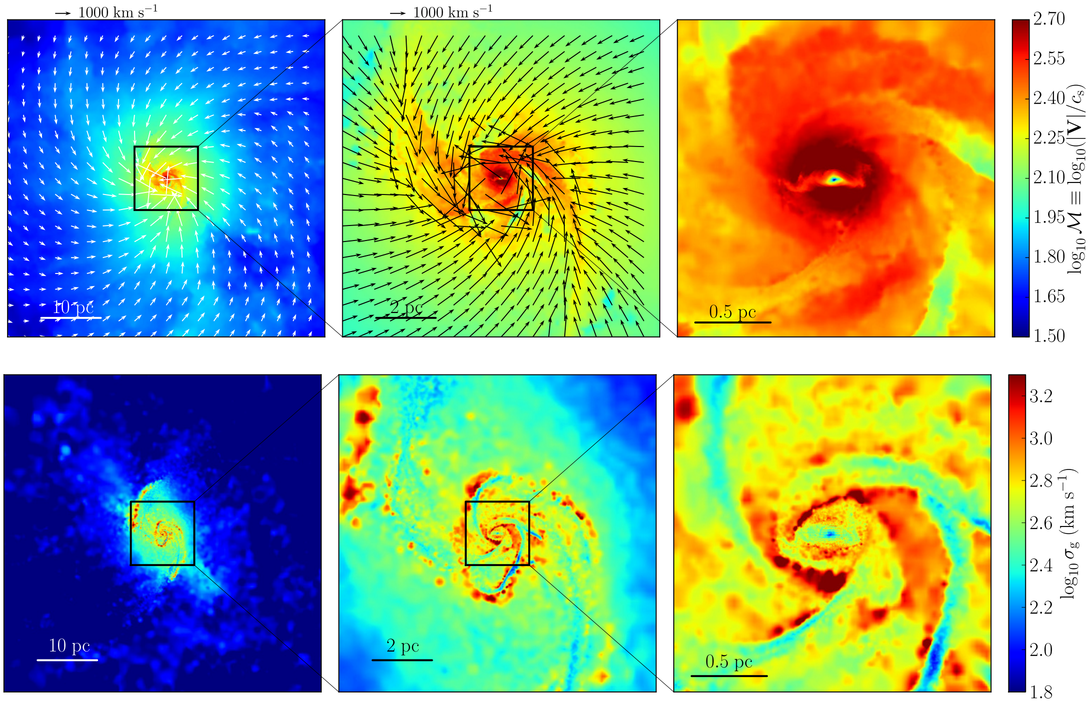

FIG. 2.— Color-coded maps at different scales of the density-weighted Mach number $\mathcal{M} \equiv |\mathbf{V}|/c_{\rm s}$ (top row) and of the density-weighted 3D gas velocity dispersion (bottom row) at time t_0 for RUN4. The first two columns of the first row show the velocity field of the gas. The flow is highly supersonic ($\mathcal{M} > 100$) and turbulent ($\sigma_{\rm g} \lesssim 1000$ km s^{-1} in the inner region), and it maintains this status till the end of the simulation.

1000 km s^{-1}, and highly supersonic, as shown by the Mach number map at $t \geq t_0$ in the top row of Figure 2 for RUN4. Such high infall velocities are the imprint of the large scale galaxy merger, which has previously initiated efficient extraction of angular momentum of the gas via gravitational torques and shocks at kpc scales, deepening the potential well of the system by pushing mass inward. The supersonic flow is thus triggered and maintained by the large scale gravitational tides. We define the Mach number $\mathcal{M}$ as the ratio between the module of the gas 3D velocity $|\mathbf{V}|$ and the local thermal sound speed $c_{\rm s}$. In the inner ~ 5 pc, $\mathcal{M} > 100$, leading to strong shocks and the development of supersonic turbulence, as shown by the 3D velocity dispersion maps in the bottom row of Figure 2. The source of turbulence kinetic energy is thus ultimately the gravitational orbital energy of the inflowing gas filaments and galaxy cores. The local velocity dispersion is measured as $\sigma_{\rm g} = \sqrt{\langle|\mathbf{V}|^2\rangle - |\langle\mathbf{V}\rangle|^2}$, where $\mathbf{V}$ is the local 3D gas velocity and the average $\langle\cdot\rangle$ is intended as SPH average on the smoothing kernel.

After the final coalescence of the two cores, the residual angular momentum in the infalling gas establishes a centrifugal barrier and leads to the formation of a self-gravitating nuclear disk with radius ~ 5 pc, as shown in the gas density projections of Figure 3. At the center, the gas accumulates in a compact, disky core that shows non-axisymmetric features such as $m = 1 - 2$ modes. This final configuration appears to be largely independent on the specific SPH implementation and cooling module adopted in the various runs (Figure 3) and is maintained till the end of the simulation. We emphasize that the nuclear disk is extremely well resolved owing to particle splitting, with nearly half a million particles.

The central disk is surrounded by filamentary rings of accreting material which exhibit substantial rotation out to several pc during the whole simulation. This is shown by the radial profile of $V_\phi/\max(\sigma_{\rm g}, |V_R|)$ in Figure 4. $V_\phi/\max(\sigma_{\rm g}, |V_R|)$ quantifies the balance between rotation and random motions by means of the ratio between the azimuthal velocity V_ϕ and the maximum between the radial velocity and the local velocity dispersion $\max(\sigma_{\rm g}, |V_R|)$. In all runs the central disk undergoes only moderate fragmentation despite being massive and self-gravitating. Indeed, the rotating disk and surrounding infalling filamentary rings are stable except along the densest spiral arms, where gravitationally bound clumps are produced via fragmentation (the stability of the nuclear region is described with detail in Section 3.2).

Figure 5 shows the time evolution of the gas mass inflow, which we interpret as a gas accretion flow onto the central disky core. The mass inflow rate is measured inside cylindrical shells with vertical thickness of 2 pc as $\dot{M} = \Delta R^{-1} \sum_j m_j V_{R,j}$, where ΔR is the radial width of the shell, m_j and $V_{R,j}$ are the mass and the radial velocity, respectively, of the j-th gas particle inside the annulus. Figure 5 shows that accretion rates of $\sim 10^4$ $M_\odot$ yr^{-1} are sustained from ~ 20 pc, where the radial motions dominate the dynamics of the gas, down to the scale of the inner compact disc[7] ~ 1 pc, where the gas is rotationally supported (see Figure 4).

The interplay between the strong shock heating associated with the highly supersonic inflow and radiative cooling is such that the nuclear disk maintains a tem-

---

[7] The outflow occurring at the final stage of the evolution in the inner region are due to the presence of a second massive core; see Section 4 for details.

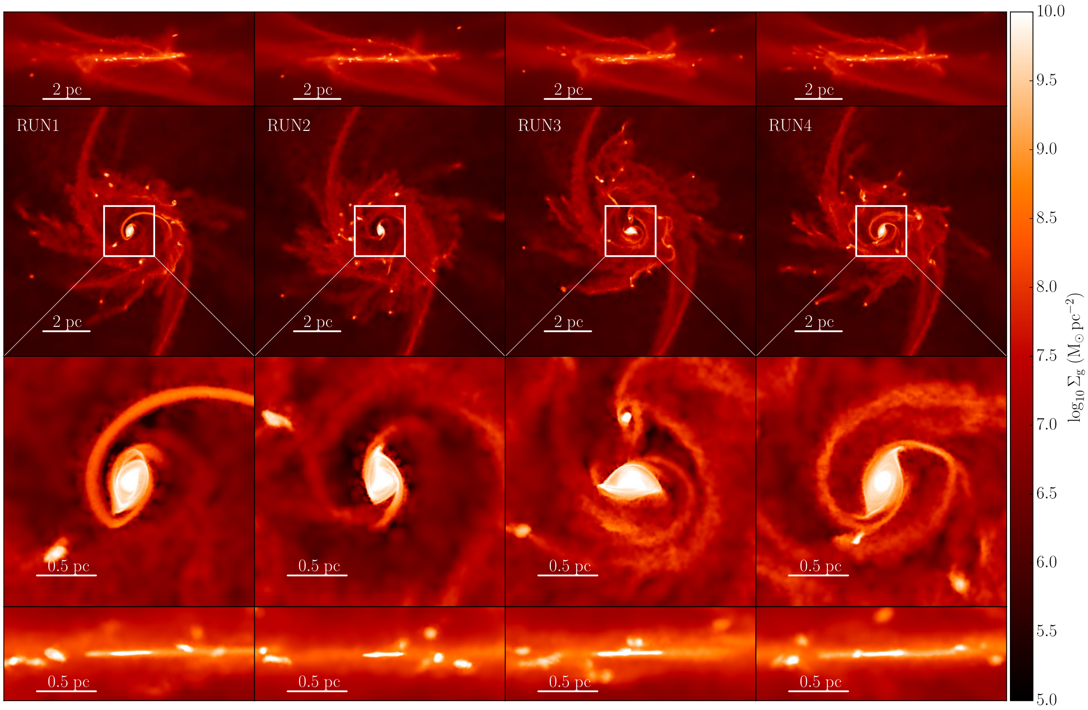

FIG. 3.— Face-on and edge-on projected gas density maps of the nuclear region of the merger at $t_0 + 5$ kyr, showing a disk-like object with radius $\lesssim 5$ pc (first and second row edge-on and face-on, respectively), and a compact inner disk-like core less than a pc in size (third and fourth row face-on and edge-on, respectively). From left to right, the columns compare RUN1, RUN2, RUN3, and RUN4, respectively, with labels according to Table 1. The larger scale disk shows some fragmentation, while the inner region is subject to strong non-axisymmetric instabilities, but negligible fragmentation. The evolution is weakly sensible to the specific SPH implementation or to the adopted model for cooling. The edge-on views confirm the highly turbulent motions with massive clouds departing from the disk midplane.

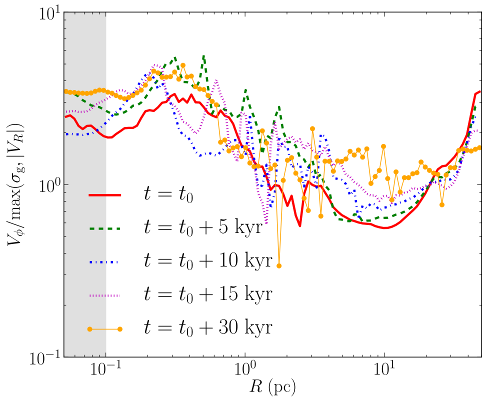

FIG. 4.— Time evolution for RUN4 of the profile of the ratio between the azimuthal velocity V_ϕ and the maximum between the radial velocity $|V_R|$ and the local 3D velocity dispersion σ_g. The gray band highlights the resolution limit given by the gravitational softening.

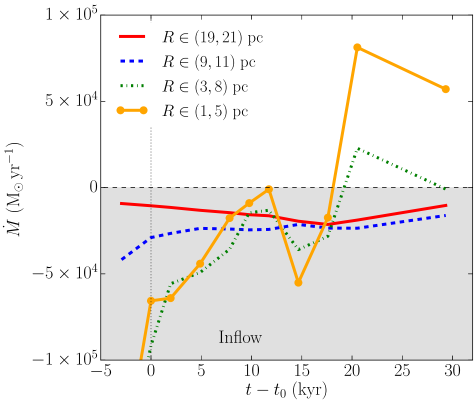

FIG. 5.— Time evolution of the gas accretion rate at different radii from the center of the merger remnant. The accretion is computed inside cylindrical shells of inner and outer radii marked in the legend and vertical thickness 2 pc (see the text for further details).

perature $\sim 3000 - 6000$ K for $t > t_0$. This is shown in Figure 6, where we compare the phase diagrams of all the runs at $t = t_0 + 5$ kyr, highlighting the nuclear disk region with black contours. The phase diagrams are very similar in the high density region occupied by the nuclear disk, regardless of the particular cooling implementation (i.e. "metal cooling" with or without the "thermal balance model") and the specific formulation of the SPH hydrodynamic force. In the absence of any cooling, an adiabatic, strong shock at the infall velocity found in the simulations (~ 1000 km s^{-1}) would induce a post-shock temperature $\gtrsim 10^6$ K. However, the presence of cooling regulates the temperature in the whole nuclear disk ($\lesssim 5$ pc) to a few thousand K (note that the cooling rate drops significantly below 10^4 K even for solar metallicity gas). We verified that the combination of high densities and very high infall velocities in the inner pc region produces a compressional heating which is roughly equal to the cooling rate by fine structure lines below 1000 K, while cooling dominates at temperatures

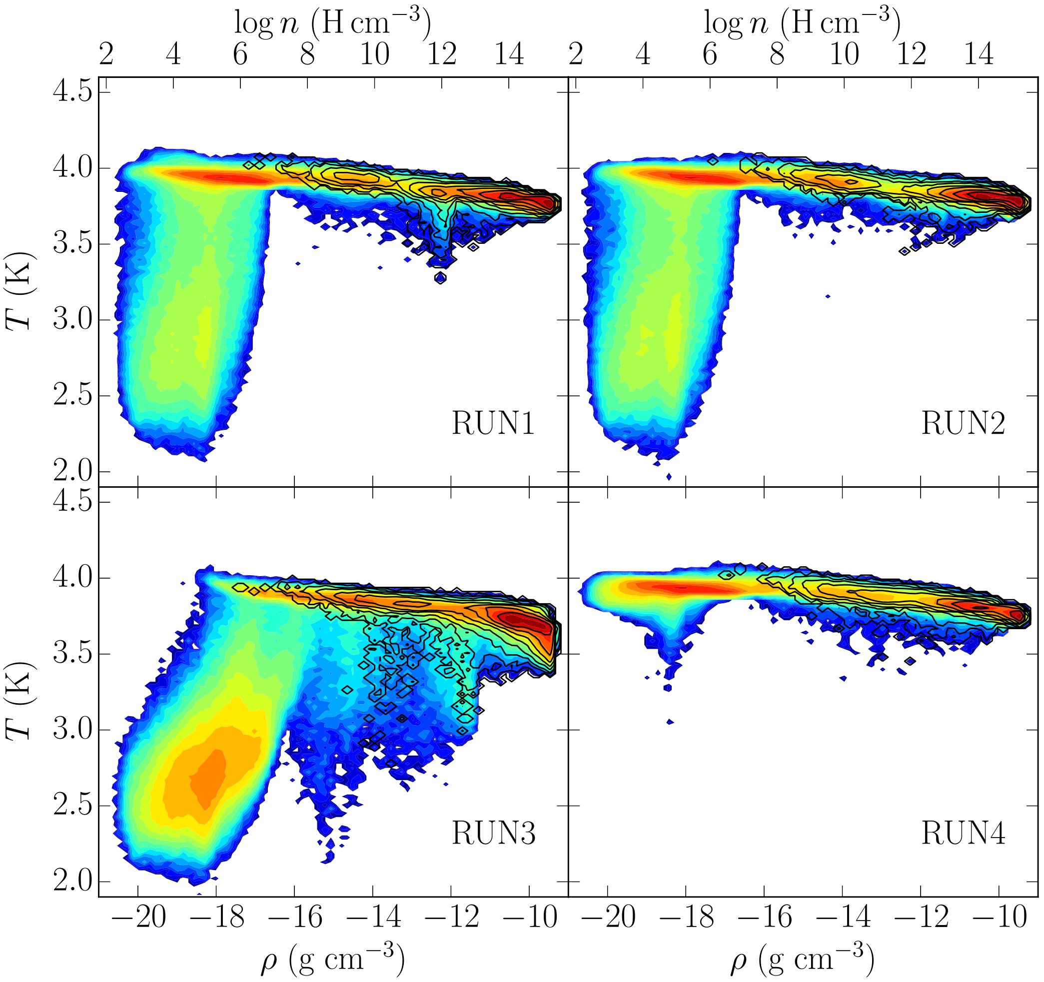

FIG. 6.— Phase diagrams of the particles within a 50 pc volume for the same runs shown in Figure 1, at $t = t_0 + 5$ kyr. From top-left to bottom-right, we show results for RUN1, RUN2, RUN3 and RUN4. Black contours are superimposed to show the particles that are located in the inner parsec region, within the inner compact disk (see text).

above 5000 K, explaining why the temperature of the core hovers around a few thousand K in the core. This is the case even without considering self-shielding effects in the optically-thick gas, which are only taken into account in the thermal balance model adopted in RUN4. The phase diagrams also show that the "metal cooling" is more effective at lowering the temperature in the low density, low Mach number regions far from the core. The latter appears less effective when GDPSH is employed (RUN4), likely because the new SPH implementation facilitates mixing between different gas phases when used in conjunction with diffusion (Wadsley et al., in preparation). Finally, cooling is enhanced at both low and intermediate densities when the thermal balance model is used (RUN3) because in such a model the cooling rate for solar metallicity gas is increased at low and intermediate densities, before absorption and re-radiation by dust suppresses cooling at high optical depths (Roškar et al. 2015). We conclude that *the warm temperature maintained by the central disk-like core seems to be a robust product of the supersonic gravitational infall established by the larger scale dynamics of a major galaxy merger such as the one simulated here.* We also note that no star formation occurs in the warm core simply because the gas never cools enough to meet the temperature condition of our star formation prescription (see Section 2.2). Instead, sporadic star formation can occur in the densest gas pockets further away from the center, where metal-line cooling is effective.

Since the the structure and thermodynamics of the inner region within ∼ 5 pc does not change appreciably among the different runs, in the following we will focus on the results of RUN4, which employs both GDSPH and diffusion. This simulation should capture better mixing as well as the multi-phase structure of the flow (Keller et al. 2014).

3.2. *A supermassive compact disk in a gravoturbulent, optically thick inflow*

Here we discuss in detail the structural properties and evolution of the central nuclear disk. Figure 7 shows the cumulative mass measured inside cylindrical shells with vertical thickness of 3 pc. Immediately after formation, the compact disk-like core forms and weighs $\lesssim 10^9\ M_\odot$, with variations of less than a factor of 2 between the different runs, and is barely 0.5 pc in size (see Figure

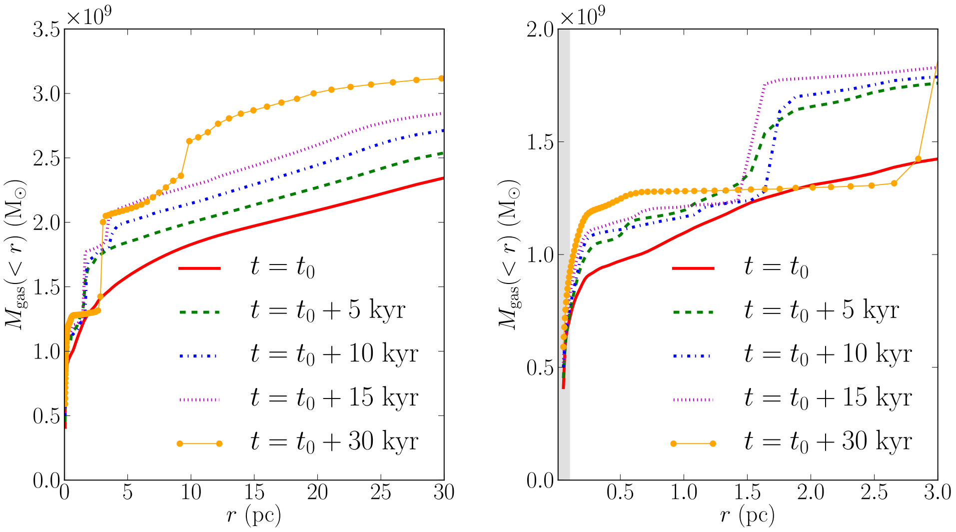

FIG. 7.— Time evolution of the enclosed gas mass inside 30 pc (left) and inside 3 pc (right), for RUN4. The reference time t_0 is the time of the coalescence of the two merging cores (see text for details). The gray stripe in the right panel highlights the resolution limit set by the gravitational softening.

3). It has a mean temperature of $\sim 4000 - 5000$ K and an aspect ratio $\sim 0.05 - 0.1$. It is born with marked non-axisymmetric structure due to its high self-gravity and mass accretion rates $\sim 10^5\ M_\odot$ yr^{-1}, which are almost an order of magnitude higher than those found in MA10 with an effective EoS. Then, the inner compact disk grows inside-out as it accretes infalling gas with increasingly higher angular momentum, reaching a mass of of $1.3 \times 10^9\ M_\odot$ within $\lesssim 1$ pc and forming a nuclear disk with an extension of $\lesssim 5$ pc after 30 kyr. Such a disk size is a factor of 10 smaller than the radius reached by the disk in MA10 soon after the merger is completed. The disk is much denser than in MA10 because the infalling gas here is much colder than with an effective EoS. The fact that the infalling gas, including the two galaxy cores, remains much colder than in MA10 allows more gas to reach the center faster, thus increasing the central baryonic mass density relative to MA10. This allows to build rapidly the central compact core, thus entering quickly a regime analogous to inside-out protostellar collapse (see e.g. Stahler & Palla 2005), in which the accretion rate is driven by gravitational accumulation onto the core. In the latter regime the inflow rate does not depend on the ambient sound speed, rather is proportional to the free-fall time on the central massive core, $\dot{M} \sim {v_{\rm ff}}^3/G$ (of course neglecting the effect of angular momentum). For a central core mass of $> 10^8\ M_\odot$ one has $v_{\rm ff} > 10^3$ km s^{-1} at a distance of 1 pc, consistent with the infall velocities we measure in the simulations. This results in $\dot{M} > 10^4\ M_\odot$ yr^{-1}, roughly consistent with the accretion rate measured at such scales. The free-fall time is $t_{\rm ff} \sim R_{\rm cl}/v_{\rm ff}$, where $R_{\rm cl}$ is the size of the infalling envelope, and with $v_{\rm ff} = 10^3$ km s^{-1} and $R_{\rm cl} = 10$ pc is as small as 10 kyr. Not surprisingly, the inner compact disk is quickly established, with a central density exceeding $\sim 10^{-10}$ g cm^{-2}, comparable to the density that in MA10 was reached only after 10^5 yr as a result of a secondary infall inside the disk triggered by spiral instabilities. The bottom line is thus is that, owing to radiative cooling, the dynamical evolution of the nuclear region is both stronger and accelerated in these new simulations relative to MA10.

The fast accretion rate on to the nuclear disk/inner core continuously maintains a high level of self-gravity (see e.g. Boley 2009 and Hayfield et al. 2011 for an analogous behavior in protostellar disks). This induces the nuclear disk to develop global spiral modes at a few parsec scale and an oval distortion within the central parsec (see Figure 3). In particular, we expect that a further increase in spatial resolution would turn the oval distortion into a strong bar-like mode since we measure $T_{\rm rot}/W \sim 0.1$ within 0.5 pc, where $T_{\rm rot}$ is the rotational kinetic energy and W the gravitational binding energy. This value is on the low side relative the condition $T_{\rm rot}/W > 0.1 - 0.3$ for bar formation in 3D rotationally flattened, self-gravitating axisymmetric configurations (Durisen & Tohline 1985; Pickett et al. 1996; Christodoulou et al. 1995). However it has been shown that the exact threshold value for stability depends on the internal structure of the rotating fluid configuration, and only a limited range of idealized equilibrium systems not subject to any mass accretion have been studied (Christodoulou et al. 1995). In addition, gravitational softening damps the growth of non-axisymmetric instabilities when their characteristic wavelength is of order the softening length (Mayer & Wadsley 2004; Debattista et al. 2006; Kaufmann et al. 2007), hence the value of $T_{\rm rot}/W$ that we measure should be interpreted as a lower limit to the value that the system might achieve.

We expect that the unresolved bar mode in the core will sustain an inflow down to $\ll 10^{-1}$ pc, as found in circumnuclear disk simulations of Choi et al. (2013). This is also in line with the general picture of bars-in-bars proposed by Begelman et al. (2006) (see also Begelman & Shlosman 2009), except for the much higher inflow rates relative to isolated protogalactic disks owing to the trigger from larger scales provided by the merger dynamics. Indeed galaxy mergers are well known to provide naturally the conditions for efficient angular momentum

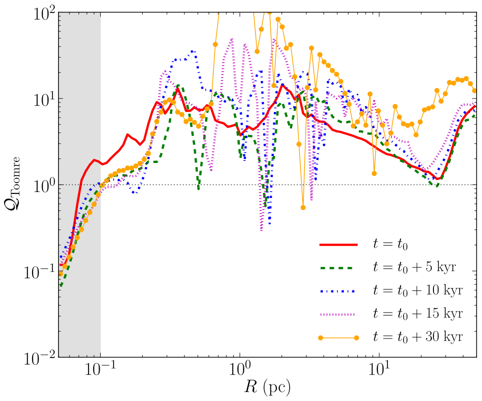

FIG. 8.— Time evolution of the Toomre parameter profile measured from RUN4 (see the adopted definition of $\mathcal{Q}_{\rm Toomre}$ in the text). Dips correspond to regions where clumps form from gas fragmentation along spiral arms, predominantly outside the inner pc region (see Figure 3 for comparison on the spatial distribution of clumps). The dotted horizontal line marks $\mathcal{Q}_{\rm Toomre} = 1$ for reference. The gray band highlights the resolution limit given by the gravitational softening.

removal owing to the combined action of strong tidal torques arising from large distortions in the mass distribution as well as large scale shocks (Kazantzidis et al. 2005). As the system will lower its rotational energy due to angular momentum transport the bar mode mode will weaken, but other instabilities can intervene at lower $T_{\rm rot}/W$, for example one-armed modes (Pickett et al. 1996). Considering the conservative case of one-armed modes, less effective than bars at shedding angular momentum, a decrease of a factor of 2 in the specific angular momentum is still expected after a few core rotations (Pickett et al. 1996).

We now examine the conditions of the self-gravitating gas flow in the central disk more closely. Fragmentation in a (laminar) rotating gas disk requires two conditions: (i) the Toomre parameter[8] $\mathcal{Q}_{\rm Toomre} < 1.5$; *and* (ii) the local cooling timescale[9] has to be shorter than the local orbital time (e.g. Lodato & Natarajan 2006; Durisen et al. 2007). Figure 8 shows the time evolution of the $\mathcal{Q}_{\rm Toomre}$ profiles of our simulation RUN4. We compute the Toomre parameter taking into account supersonic turbulence in the gas as $\mathcal{Q}_{\rm Toomre} = v_{\rm turb}\kappa/(\pi G \Sigma)$, where κ is the epicyclic frequency, Σ is the gas surface density, and $v_{\rm turb} = \sqrt{\sigma_{\rm g}^2 + c_{\rm s}^2}$. Here, $\sigma_{\rm g}$ is the 3D velocity dispersion of the gas (measured as in Section 3.1) and $c_{\rm s}$ the thermal sound speed. The $\mathcal{Q}_{\rm Toomre}$ stability threshold is satisfied in most of the nuclear disk after a few thousand year since its emergence, except for some minima below 1.5 typically arising at $1 \lesssim R/{\rm pc} \lesssim 5$ along spiral arms in the nuclear disk (see Figure 3), which indeed is where gaseous clumps are observed to form via fragmentation. Supersonic turbulence provides support against gravitational collapse, both directly (kinematically) and via heating through small scale shocks as typically observed in star formation simulations (Padoan et al. 1997). Therefore, turbulence influences also the cooling time criterion increasing the effective cooling time. Although we can not probe directly the non-linear coupling between turbulence heating and radiative cooling, we argue that supersonic turbulence plays an important role in regulating the temperature of the nuclear region, since $\sigma_{\rm g} \gg c_{\rm s}$ and the runs appear rather insensitive to the details of the cooling within a few parsecs. The key role of turbulence in regulating disk fragmentation was noted also by Choi et al. (2013) and Latif et al. (2013).

We do not include radiative transfer in our simulations, although the "thermal balance model" can be viewed as an approximate way to introduce radiative transfer effects. Nevertheless, in order to support further our results we can show analytically that the intrinsic optically-thick nature of the nuclear disk would suppress fragmentation in the inner parsec. Indeed, the typical optical depth in the disky core within 1 pc is $\tau_{\rm es} \sim N_{\rm g}\sigma_{\rm T} \sim 10^4$, where $\sigma_{\rm T} \simeq 6.65 \times 10^{-25}$ cm^2 is the Thomson scattering cross section and $N_{\rm g} \sim 10^{29}$ H cm^{-2} is the mean gas column density within 0.5 pc (corresponding to a surface density $\sim 10^9\ M_\odot$ pc^{-2}). Note that $N_{\rm g}$ varies by an order of magnitude above an below the quoted value in the very inner region and in the low density gaps between spiral arms and rings, respectively. At the temperature of the core, which is $\sim 3000 - 5000$ K, the opacity due to electron capture by H$^-$ might indeed be up to ten times higher than $\sigma_{\rm T}$ for the densities in the inner parsec ($\sim 10^{-10}$ g cm^{-3}), making our estimate of the optical depth conservative. Finally, the gas would then cool on the photon diffusion timescale $t_{\rm diff} \sim H\tau_{\rm es}/c \sim 3000$ yr, where c is the speed of light and $H \sim 0.1R \sim 0.1$ pc is the vertical scale height of the disk. This timescale is much longer than the orbital time, which is ~ 500 yr for $R < 1$ pc, meaning that no fragmentation should occur. Note for comparison that Ferrara et al. (2013) determined a similar diffusion timescale[10], but they compared it with a free-fall time two orders of magnitude longer then the orbital time we find in our simulation, concluding that fragmentation would be ubiquitous in the post-merger core. This is because they underestimated the central density of the disk due to their isochoric assumption. In other words, they explored the effect of radiative cooling without considering how both cooling and shocks in the infalling gas would modify the density structure of the disk in the first place.

The conditions in the disk outside $\sim 1 - 2$ pc change because the average velocity dispersion decreases, the orbital time increases and fragmentation can take place (see upper panel of Figure 3). The co-existence of an inner region stable to fragmentation and an outer region unstable to fragmentation is reminiscent of massive self-

---

[8] The conventional stability threshold $\mathcal{Q}_{\rm Toomre} > 1$ is derived using linear perturbation theory for an infinitesimally thin, axisymmetric disk. However, higher order perturbation theory and numerical simulations in 2D and 3D favor the phenomenological threshold $\mathcal{Q}_{\rm Toomre} > 1.5$ in more general conditions (e.g. Durisen et al. 2007).

[9] The cooling timescale condition is well established in simulations of self-gravitating gas disks (e.g. self-gravitating protoplanetary disks; Gammie 2001; Rice et al. 2003; Mayer et al. 2005; Meru & Bate 2011a; Helled et al. 2013). However, the exact value of the ratio between the cooling time and the orbital time is still a matter of debate since proof of convergence as a function of resolution and cooling law implementation has not been achieved yet (see e.g. Meru & Bate 2011b; Helled et al. 2013). Therefore, we will be intentionally conservative and assume that a cooling time equal to the orbital time is a sufficient second condition for fragmentation.

[10] They assumed a lower optical depth but a higher value for H.

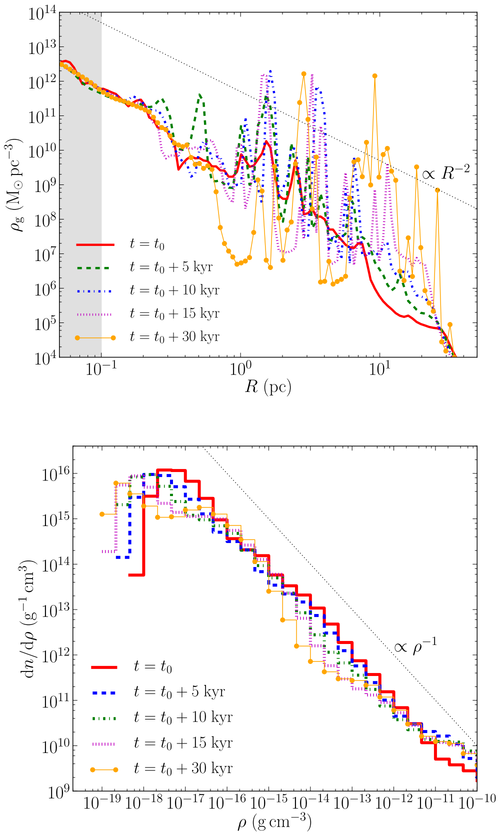

Fig. 9.— Time evolution of the gas density radial profile for RUN4 (top). The dotted line show the $\propto R^{-2}$ scaling. Time evolution of the gas density PDF (bottom). The dotted line show the $\propto \rho^{-1}$ scaling. The density PDF is computed inside the cylindrical region with radius of 10 pc and with height of 5 pc centered on the nuclear disk.

gravitating protostellar and proto-planetary disks (e.g. Boley et al. 2010; Helled et al. 2013). Large scale spiral instabilities are present, and massive clumps in the range $10^6 - 10^8$ $M_\odot$ form along the arms. Most of them appear to be only marginally bound, being dispersed and reforming on a few orbital timescales. The typical temperature of the clumps is > 2000 K, hence no conventional star formation can take place inside them. Clumps that are dense enough to survive tides will spiral towards the center as a result of dynamical friction on a few orbital times (Noguchi 1998; Bournaud et al. 2009; Fiacconi et al. 2013).

The gravoturbulent nature of the flow is reflected also by the density profile and the density probability distribution function (PDF) acquired by the gas, shown respectively in the upper and lower panel of Figure 9 at different times. The mass density profile achieved in the inner 10 pc as a result of the coincident action of self-gravity, turbulence and thermodynamics is close to $\propto R^{-2}$. Such a density slope has been shown to favor central collapse relative to fragmentation in the envelope for gravoturbulent molecular clouds, a condition that is often invoked to form a massive star rather than a cluster of lower mass stars (Girichidis et al. 2011). This qualitatively corresponds to the behavior observed in our simulations. The density PDF of the gas in the nuclear region carries the imprint of supersonic turbulence in presence of self-gravity. It exhibits a slope close to -1, in nice agreement with the simulations of Choi et al. (2013), as opposed to the slope of ~ -2 typically found before self-gravitating collapse ensues in simulations of supersonic turbulence in the ISM (e.g. Scalo et al. 1998; Kritsuk et al. 2011). It has been suggested that a slope ~ -1 may result from dynamically important angular momentum (Kritsuk et al. 2011), and indeed we verified that such region corresponds to the region where the nuclear disk assembles, again in agreement with Choi et al. (2013). However, our PDF cannot be simply described by a log-normal distribution function at low densities as in Choi et al. (2013). We argue that the excess amplitude relative to a log-normal PDF at densities $\rho < 10^{-15}$ g cm^{-3} is due to the large amount of infalling gas exterior to the inner compact disky-core. Presumably less gas is present at large radii and lower density in Choi et al. (2013) since these authors model an isolated protogalaxy rather than a galaxy merger. Our interpretation is supported by the fact that the low density tail tends to become slightly lower with time as less infalling continues to reach the inner few parsecs.

4. POSSIBLE EVOLUTIONARY PATHWAYS OF THE INNER COMPACT CORE

We stop our simulations between 30 and 50 kyr after t_0 because we reached prohibitively small timesteps and the central inflow is saturated at the resolution scale set by the gravitational softening. Although we cannot ascertain the ultimate fate of the gas below the resolution limit, the simulations provide a wealth of information that allows us to delineate possible scenarios to form a massive BH seed given the final conditions inside the pc-size compact disky core present in all of our runs. Hereafter we will refer to the final conditions of RUN4 described in the previous section.

The mass accretion rate has been shown to be the key factor deciding whether a SMS will form or the system will remain in a protostar-like phase, on the Hayashi track, and produce a quasi-star directly in $\sim 10^7$ yr (Begelman 2010; Schleicher et al. 2013; Choi et al. 2013). Schleicher et al. (2013) assumed an adiabatic spherical steady contraction without rotation, and found that the critical mass necessary for forming a SMS increases with the gas accretion rate as 3.6×10^8 $\dot{m}^3$ $M_\odot$, where $\dot{m}$ is the mass accretion rate normalized such that $\dot{m} = 1$ for an accretion rate $\dot{M} = 1$ $M_\odot$ yr^{-1}. The accretion rates found in all our runs in the inner compact disk are typically $> 10^4$ $M_\odot$ yr^{-1}, which would yield a critical mass a few orders of magnitude higher than that of the compact core. However, both the QS or the SMS would form below our resolution scale set by the gravitational softening (Hosokawa et al. 2012; Hosokawa et al. 2013), hence the quoted accretion rate is not necessarily representative of the gas flow at those scales.

There is, however, another critical condition that may become relevant on much shorter timescales. This is the condition for global relativistic radial collapse. In this case the mass accretion rate does not appear directly in the equations but may still be instrumental to achieve the critical mass in the first place. Numerical general-relativistic (GR) simulations show that the instability for a rotating fluid configuration (a polytrope with $\gamma = 4/3$) is reached for: (i) a *compactness threshold* $R < 640\,GM/c^2$, and (ii) a dimensionless spin parameter $q \equiv cJ/(GM^2) < 0.97$ (Baumgarte & Shapiro 1999; Shibata & Shapiro 2002; Saijo & Hawke 2009; Reisswig et al. 2013), where M is the mass and J the total angular momentum of the cloud. This has been shown to apply to both uniformly rotating and differentially rotating clouds. The inner disky core in our simulations reaches quickly $\sim 10^9\ M_\odot$ within 0.2 pc and it has typically $q \sim 15$. On the other hand, the compactness threshold for $M \sim 10^9\ M_\odot$ is $R \sim 0.03$, which is ~ 6 times smaller than the characteristic radius of 0.2 pc. Therefore, the disky core would seem to be stable to the general relativistic (GR) radial instability. However, as we outlined in section 3.2 we expect that it will become bar-unstable, which would transport angular momentum outward and cause further contraction, possibly pushing it closer to the verge of the instability.

An alternative, perhaps more effective way to discuss stability is to rely on another result of relativistic simulations. Indeed, starting from 2D and 3D rotating polytropic clouds with masses exceeding $10^6\ M_\odot$, these simulations show that $T_{\rm rot}/W \sim 0.01$ or lower is a sufficient condition to bring the cloud to the radial collapse stage under a small initial perturbation [11] (Shibata & Shapiro 2002). This is of course a phenomenological criterion, and may change with a different EoS, but offers a useful guideline. If we adopt the latter condition, the angular momentum in the inner compact disk has to decrease substantially to enter radial collapse since $T_{\rm rot}/W \sim 0.1$ at radii $\lesssim 0.5$ pc. Drawing from the calculations of bar-unstable protostellar clouds, which can apply here since the eventual subsequent contraction will mostly be in the newtonian regime, one expects a decrease of the specific angular momentum j by a factor of 2 over a few dynamical times, (e.g. Pickett et al. 1996), i.e. over $< 10^4$ yr. At the same time, the mass of the system can grow up to a factor of ~ 2, if accretion rates $> 10^4\ M_\odot$/yr are sustained by the bar down to scales < 0.5 pc for $\sim 10^5$ yr. Therefore, $T_{\rm rot}/W$ would decrease by almost an order of magnitude at fixed radius on relatively short timescales (since it scales as $T_{\rm rot}/W \propto j^2 M^{-1} R^{-1}$), reaching the critical threshold for radial collapse.

Although we cannot reach firm quantitative conclusions, our simple estimates suggest the direct massive BH formation via GR radial instability might occur if gas inflow rates as those measured un our simulations are maintained down to scales close to our resolution limit for a time only slightly longer than we could probe here. Since up to $\sim 60 - 90\%$ of the progenitor mass can be retained during GR radial collapse (e.g. Saijo & Hawke 2009; Reisswig et al. 2013), the emerging BH seed would have a mass roughly between 10^8 and $10^9\ M_\odot$, namely in the SMBH mass range and, most importantly, already close the mass inferred for the SMBHs powering the high-z QSOs.

Finally, another intriguing aspect of the simulations is that, at least in RUN4, a binary system of two compact disk-like cores eventually arises. This is shown in the gas density projection in Figure 10 at time $t_0 + 30$ kyr. The secondary clump forms after about 25 kyr from fragmentation along a prominent spiral arm. It has a mass only 5 times smaller than that of the primary, which is $\gtrsim 10^9\ M_\odot$, and a comparable size. Its appearance is also highlighted by the bump at $r \sim 3$ pc in the cumulative mass profile shown in Figure 7. It also appears as an "outflow" feature in Figure 5 at $t > 20$ kyr as it orbits the center of mass of the system. At this point the two massive clumps may begin to compete for gas accretion, analogously to the case of multiple protostellar cores in star forming regions. However, we argue that this should mildly affect the onset of global relativistic collapse.

Simulations of massive black hole binaries which evolve in a similar environment suggest that the separations of the two objects will be reduced to < 1 pc by dynamical friction against the background in less than 1 Myr (e.g. Mayer 2013). If a massive BH seed forms in the meantime at the center of both objects one may speculate that a merger between two massive BHs will occur, leading to a burst of gravitational waves (GWs). The frequency of such signal may vary a lot depending on the mass ratio of the BH seeds and on their orbital eccentricity. Assuming conservatively the BH seed mass found in models of direct collapse that posit a quasi-star stage ($\sim 10^3 - 10^4\ M_\odot$; Dotan et al. 2011), the signal could be detected by the *eLISA* probe (Amaro-Seoane et al. 2013). Alternatively, if a secondary massive clump is not formed during the stage studied here, the central core might still fragment later during the relativistic phase of the collapse. Indeed in the latter phase a relativistic bar instability could be triggered, which may lead to fission into two or more massive BHs. This was demonstrated by Reisswig et al. (2013) with 3D relativistic hydrodynamics simulations starting from polytropic, marginally unstable, differentially rotating SMS.

We caution that the arguments outlined in this section do not address the thermodynamical evolution during the final collapse stage, which will not necessarily produce a simple polytropic configuration. Hence, only by evolving the end state of our system with a numerical GR code along with a proper treatment of radiation hydrodynamics we will be able to ascertain whether or not direct formation of a SMBH is possible.

5. DISCUSSION AND CAVEATS

We have presented simulations with greater physical realism than those in our previous work (MA10). Our results lend stronger support to the scenario presented in MA10 and Bonoli et al. (2014), in which the SMBHs powering the bright QSOs at $z > 5$ form by direct collapse driven by multi-scale gas inflows in major mergers between the most massive galaxies already present at $z \sim 8 - 10$. A key difference with all other direct collapse models presented in the literature is that here we do not require metal-poor gas or dissociation of H_2 to suppress cooling, rather radiative cooling itself fosters the rapid formation of a compact and dense nuclear disk. Fragmentation can happen, but the inner pc-scale core is stabilized by shock heating, gravoturbulence and the high optical depth of the gas. The gas collapses faster and it reaches higher densities relative to MA10 simulations, which adopted and effective EoS rather than in-

---

[11] Recently Montero et al. (2012) have shown that if nuclear burning via the CNO cycle begins at the very center it can induce a shock wave that deflagrates the protostar but so far this has been obtained only in 2D simulations, and for systems with mass $10^5 - 10^6\ M_\odot$.

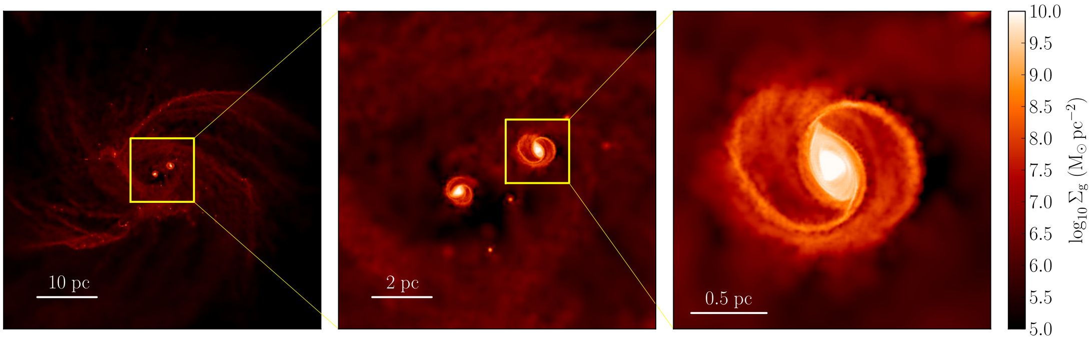

FIG. 10.— Projected gas density maps of a snapshot at time $t = t_0 + 30$ kyr, showing the formation of a secondary massive clump as a result of fragmentation (see text). Each panel, from left to right, shows a different scale. In particular, a zoom-in on the primary core is shown on the right.

corporating cooling. While the subsequent transport of angular momentum below 0.1 pc will have to be explored further, we have argued, based on simple analytical arguments, that rapid removal of angular momentum by non-axisymmetric instabilities might continue below the resolution of our simulations. Most importantly, we have highlighted the possibility that the compact disky-core, which encompasses a mass of nearly 10^9 $M_\odot$, may reach the condition for the onset of the GR radial instability by collapsing just an extra decade in radius.

When such a condition is reached, GR simulations tell us that the inner core would leave behind a BH seed exceeding half of its mass. Being really conservative, we can posit that the BH seed will have a mass of 10^8 $M_\odot$. With such a seed, little accretion will be needed to reach rapidly the BH masses inferred for high-z QSOs. Indeed, the BH would take $\tau_{\rm acc} \approx 10^8\, f_{\rm edd}^{-1}\, (\log(M_{\bullet,\rm f}/M_{\bullet,\rm i})/\log(10))$ yr to grow from $M_{\bullet,\rm i}$ to $M_{\bullet,\rm f} = 10 M_{\bullet,\rm i}$. If we assume an Eddington ratio $f_{\rm edd} \approx 0.3$ all the time, the typical value obtained in radiative simulations of gas accretion onto BH seeds (Milosavljević et al. 2009), we obtain $\tau_{\rm acc} \approx 3 \times 10^8$ yr, comfortably shorter than the time elapsed to $z = 6-7$ in a Λ-CDM cosmology. Note that the Eddington accretion rate would change from ~ 2 to ~ 20 $M_\odot$ yr^{-1} between $\sim 10^8$ and $\sim 10^9$ $M_\odot$. The gas reservoir in the inner 100 pc of the merger remnant is $\sim 3 \times 10^9$ $M_\odot$ and undergoes infall, hence there is plenty of gas to accrete from, a point already emphasized in MA10.

The thermodynamics of the gas will change drastically after a BH seed forms at the center. If the seed emerges already with the mass of a SMBH, any energetic feedback from subsequent accretion, although it can limit its further growth is of little relevance for the origin of the high-z QSOs. Instead, such an early strong feedback mode might have implications for galaxy formation as it could affect gas cooling in nearby halos (see e.g. Dijkstra et al. 2006, 2014) as well as stifle bulge formation in the host galaxy. Feedback can surely have a more important impact on the growth of lighter seeds in the range $10^3 - 10^5$ $M_\odot$. However, we have shown that the photon diffusion timescale is longer than the orbital time below parsec scales and thus longer than the free-fall timescale. This means that radiation generated by the accretion process will be trapped in the very inner region on characteristic timescale of the inflow itself. Therefore, the luminosity produced by accretion on to the BH can be integrated over the time interval Δt to yield an estimate of the energy deposited in the inner region, say roughly below 1 pc. Such integrated energy release is $E_{\rm fb} \sim \epsilon_{\rm fb}\, L_\bullet\, \Delta t$, where $\epsilon_{\rm fb}$ parametrizes uncertainties on the coupling between gas and radiation, and $L_\bullet$ is the accretion luminosity on to the BH, which we roughly estimate as the Eddington luminosity for a 10^5 $M_\odot$ BH. Note $\epsilon_{\rm fb}$ is highly uncertain, but we take it ~ 0.01 as usually done in effective models of BH feedback (see e.g. Di Matteo et al. 2005; Van Wassenhove et al. 2014). Over a timescale of $\Delta t = 10^5$ yr we obtain $E_{\rm fb} \sim 4 \times 10^{53}$ erg. This should be compared to the binding energy of the gas, which is $E_{\rm b} \approx GM_{\rm disk}^2/R_{\rm disk} \sim 10^{59}$ erg assuming $R_{\rm disk} = 1$ pc and $M_{\rm disk} = 10^9$ $M_\odot$. Therefore feedback should not be strong enough to stifle accretion even if the BH seed accretes up to five orders of magnitude above Eddington.

Based on the scenario outlined in the previous section, the massive BH formation process takes less than 10^5 yr following the merger if GR radial collapse occurs. This implies that bright QSOs with SMBHs $\sim 10^8 - 10^9$ $M_\odot$ can appear essentially as soon as galaxies massive enough to host sufficiently abundant gas reservoirs and multi-scale scale inflows arise. In MA10 and Bonoli et al. (2014) we studied the dependence of the gas inflow on galaxy mass (but fixed gas fraction), finding that a major merger between galaxies having virial mass $> 10^{11}$ $M_\odot$ is a necessary condition for the inflow to produce a central collapse. We can estimate roughly how the inflow rate should scale with virial mass by assuming that the inflow occurs in a self-similar way. Thus, the inflow rate would scale as $\dot{M} \sim v_{\rm ff}^3/G \propto M_{\rm vir}$, since $v_{\rm ff} \sim V_{\rm circ} \propto M_{\rm vir}^{1/3}$, where $V_{\rm circ}$ is the virial circular velocity of the halo (Mo et al. 1998). In our simulations $\dot{M} \sim 10^4$ $M_\odot$ yr^{-1} and we expect $\dot{M} \sim 10^3$ $M_\odot$ yr^{-1} in galaxies ten times lower in mass. This means that the mass accumulated at the center would be an order of magnitude lower on the same timescale and the inner compact disky-core would not be able to reach the onset of the GR radial instability, unless accretion is sustained at least ten times longer ($> 10^5$ yr). However, this longer timescale increases the chances that the inflow rate would decrease significantly before reaching the radial instability conditions and opens the path to the formation of an SMS and, eventually, of a QS. In this case the scenario would be identical to that adopted in Bonoli et al. (2014), and the seed that would form has a mass in the range $10^3 - 10^5\ M_\odot$ rather than $\gtrsim 10^8\ M_\odot$. We calculated that about 200 major mergers between halos above $10^{11}\ M_\odot$ take place at $z > 6$ in the *Millennium Simulation*, whose volume is $500^3\ h^{-3}\,\mathrm{Mpc}^3$. If all these events lead to a SMBH-like seed via GR radial instability with mass $> 10^8\ M_\odot$, the number density of quasars would be $\sim 1.6 \times 10^{-6}\ h^3\,\mathrm{Mpc}^{-3}$. If we further assume that only halos above $10^{12}\ M_\odot$ produce seeds $> 10^8\ M_\odot$ because of the scaling argument for $\dot{M}$ outlined above, then the number of events drops to only about 10, corresponding to a number density $\sim 3 \times 10^{-8}$ Mpc^{-3} Interestingly, even in this latter case the predicted number density is still high enough to explain all the $z > 6$ QSOs, which have an estimated number density of $\sim 10^{-8}\ \mathrm{Mpc}^{-3}$ (Willott et al. 2010; Treister et al. 2013). We note however that this discussion is based on the scaling arguments derived above, which do not necessarily apply to major mergers. Therefore, the dependence on galaxy mass will have to be revisited exploring a wider range of conditions for galaxy mergers, possibly drawn from cosmological simulations.

Some shortcomings are also present in the simulations themselves. A major limitation is the nature of the adopted initial conditions. They are still based on binary mergers rather than on mergers drawn from cosmological simulations. However, a vast body of literature has recently shown that the most massive high-z galaxies have turbulent, clumpy disks (e.g. Tacconi et al. 2010; Ceverino et al. 2012; Förster Schreiber et al. 2014; but see also Fiacconi et al. 2015 for lower mass galaxies). These conditions might favor prominent gas inflows during the mergers since non-axisymmetric torquing of the gas will be aided by gas turbulence. This could lead to higher central gas densities in the galaxy cores prior to merging. Clumps formed at galactic disk scales would rain down to the center as the merger comes to completion due to dynamical friction, ultimately aiding the build up of central self-gravitating mass. However, since angular momentum dissipation is crucial, changing the orbit, geometry and properties of the flow by merging misaligned galaxies or galaxies with a more turbulent ISM may change some of the detailed properties of the inflow. This in turn may affect the properties of the compact disk-like core emerging at the center of the remnant. However, it is reassuring that recent studies of the effect of merger geometry on gas inflows down to ~ 10 pc find very weak variations (Capelo et al. 2014).

Another limitation is that the simulations start from merging galaxies that had no star formation happening at scales below 100 pc because of the initial lack of resolution (i.e. before particle splitting is applied). Pre-existing star formation could have stirred the gas in the galaxy cores via stellar feedback prior to the merger, leading to a more fluffy gas distribution at the center of the remnant. We made an attempt to explore this effect by imposing instantaneous feedback in the central 100 pc region after particle splitting is applied. To achieve this we force all star particle located within this volume, each of them sampling a Kroupa IMF, to have an age of about 6 Myr, so that massive stars readily go off as SN Type II immediately nearly everywhere after the beginning of the simulation. This was done in RUN4 and the resulting effect on hydrodynamics was found to be minor. The gas closest to the sites of the explosions is heated and remains hot for the whole duration of the simulation[12], but most of the gas is not directly affected and keeps cooling and flowing inward supersonically.

A potentially more important shortcoming is the lack of feedback provided by the accretion luminosity of the central clump (see Schleicher et al. 2013), which is the analog of protostellar accretion luminosity in star formation (Krumholz et al. 2006). Drawing from the example of massive stars, radiation pressure associated with accretion luminosity may eventually stop the inflow and generate outflows. The largest accretion luminosity is generated in the inner region of the nuclear disk, where the potential well deepens the most. At ~ 1 pc the core generates $L_{\mathrm{acc}} = GM_{\mathrm{disk}}\dot{M}/R_{\mathrm{disk}} \sim 3 \times 10^{46}$ erg s^{-1} (using $R_{\mathrm{disk}} = 1$ pc, $M_{\mathrm{disk}} = 10^9\ M_\odot$, $\dot{M} = 10^4\ M_\odot\ \mathrm{yr}^{-1}$; all these quantities evolve during the simulations but by less than an order of magnitude). This is smaller than the Eddington luminosity, $L_{\mathrm{edd}} \sim 1.3 \times 10^{47}$ erg s^{-1}, for an object with mass M_{disk}, suggesting that radiation pressure should not stop the inflow. Nonetheless, during the contraction required in order to reach the stage of the GR radial instability, L_{acc} might increase and become comparable to L_{edd}. At the same, the gas optical depth will increase further and might lead to super-Eddington, radiatively inefficient accretion as mentioned above (Madau et al. 2014). Therefore, the dynamical effect of the accretion radiation has to be probed further in order to understand its impact.

The environment in the nuclear disk is highly optically thick, hence detection of the initial stages of this SMBH formation process in the electromagnetic domain may be difficult. At a temperature of a few thousand K, the inner disk will shine blackbody photons at visible wavelengths into the dense infalling gas, which will re-radiate in the infrared, resembling a protostellar envelope. The outgoing luminosity might be much less than the accretion luminosity estimated above due to absorption by the larger scale gas-rich, dusty nuclear environment. While the *James Webb Space Telescope* might be a good candidate instrument to detect the infrared signal, extinction and possibly re-emission to even longer wavelengths by the envelope of the nuclear disk might make it more accessible by the *Atacama Large Millimeter Array.* The detection of a powerful nuclear FIR source associated with highly supersonic gas infall velocities revealed with high resolution spectroscopy could represent the "smoking gun" signature of a direct collapse event. Yet the spectroscopic accuracy required to detect strong velocity gradients at scales of hundreds of parsecs at $z > 4 - 5$ is hardly attainable with current or upcoming instruments.

Ultimately GWs will provide the cleanest signature of the flavor of direct collapse that we are proposing, namely a "cold direct collapse" into a SMBH driven by the radial instability. Strong GW bursts with a characteristic wave-form and amplitude do indeed arise during the radial collapse phase of (axisymmetric) supermassive objects in numerical GR simulations, having a frequency $10^{-4} - 10^{-1}$ Hz that falls within the expected *eLISA* band (Saijo & Hawke 2009). Furthermore, if binaries of

---

[12] The blastwave feedback scheme by construction enforces shut-off of radiative cooling for a timescale of $5 - 20$ Myr (Stinson et al. 2006).

SMBHs arise commonly "at birth" in this scenario, as our simulations suggest, one would detect two relatively low frequency bursts of GWs as the two SMBHs form via the relativistic collapse, followed by an even lower frequency signal as the two holes spiral-in towards coalescence a few Myr later (Mayer 2013). A detailed study of the expected overall pattern and features of the composite signal, involving considerations of the timescale that separates birth and coalescence of the two SMBHs, with the aim at assessing detectability by the planned *eLISA* interferometer, warrants investigation in future work.

We are grateful for stimulating and helpful discussions with Piero Madau, Monica Colpi, Mitch Begelman, Francesco Haardt, Ralph Pudritz, Chris McKee, Mike Fall, Elena M. Rossi, Melvyn Davies, Milos Milosavljevic, Kevin Schawinski, Luciano Rezzolla and Stephan Reisswig. LM, DF and SB acknowledge hospitality to KITP, where crucial phases of this work were conceived and carried out in two different KITP programs in 2013 and 2014 ("A Universe of Black Holes" and "Feedback: Gravity's Loyal Opposition").

REFERENCES

Agertz, O., Teyssier, R., & Moore, B. 2009, MNRAS, 397, L64
Alvarez, M. A., Wise, J. H., & Abel, T. 2009, ApJ, 701, L133
Amaro-Seoane, P., Aoudia, S., Babak, S., et al. 2013, GW Notes, Vol. 6, p. 4-110, 6, 4
Baumgarte, T. W., & Shapiro, S. L. 1999, ApJ, 526, 941
Begelman, M. C., Volonteri, M., & Rees, M. J. 2006, MNRAS, 370, 289
Begelman, M. C., Rossi, E. M., & Armitage, P. J. 2008, MNRAS, 387, 1649
Begelman, M. C., & Shlosman, I. 2009, ApJ, 702, L5
Begelman, M. C. 2010, MNRAS, 402, 673
Begelman, M. C. 2012, ApJ, 749, L3
Belkus, H., Van Bever, J., & Vanbeveren, D. 2007, ApJ, 659, 1576
Boley, A. C. 2009, ApJ, 695, L53
Boley, A. C., Hayfield, T., Mayer, L., & Durisen, R. H. 2010, Icarus, 207, 509
Bonoli, S., Mayer, L., & Callegari, S. 2014, MNRAS, 437, 1576
Bournaud, F., Elmegreen, B. G., & Martig, M. 2009, ApJ, 707, L1
Bournaud, F., Juneau, S., Le Floc'h, E., et al. 2012, ApJ, 757, 81
Capelo, P. R., Volonteri, M., Dotti, M., et al. 2014, arXiv:1409.0004
Ceverino, D., Dekel, A., Mandelker, N., et al. 2012, MNRAS, 420, 3490
Choi, J.-H., Shlosman, I., & Begelman, M. C. 2013, ApJ, 774, 149
Christodoulou, D.M., Shlosman, I, & Tohline, J.E. 1995, ApJ, 443, 551
Davies, M. B., Miller, M. C., & Bellovary, J. M. 2011, ApJ, 740, L42
Devecchi, B., & Volonteri, M. 2009, ApJ, 694, 302
Devecchi, B., Volonteri, M., Rossi, E. M., Colpi, M., & Portegies Zwart, S. 2012, MNRAS, 421, 1465
Debattista, V. P., Mayer, L., Carollo, C. M., et al. 2006, ApJ, 645, 209
di Matteo, T., Springel, V., & Hernquist, L., 2005, Nature, 433, 604
di Matteo, Khandai, N, de Graf, C. , Feng, Y., Croft, R.A.C., Lopez, J., & Springel, V., 2012, ApJ, 745, L29
Dijkstra, M., Haiman, Z., & Spaans, M. 2006, ApJ, 649, 14
Dijkstra, M., Ferrara, A., & Mesinger, A. 2014, MNRAS, 442, 2036
Dotan, C., Rossi, E. M., & Shaviv, N. J. 2011, MNRAS, 417, 3035
Durisen, R. H., & Tohline, J. E. 1985, Protostars and Planets II, 534
Durisen, R. H., Boss, A. P., Mayer, L., et al. 2007, Protostars and Planets V, 607
Fan, X., Narayanan, V. K., Lupton, R. H., et al. 2001, AJ, 122, 2833
Ferrara, A., & Loeb, A. 2013, MNRAS, 431, 2826
Ferrara, A., Haardt, F., & Salvaterra, R. 2013, MNRAS, 434, 2600
Fiacconi, D., Mayer, L., Roškar, R., & Colpi, M. 2013, ApJ, 777, L14
Fiacconi, D., Feldmann, R., & Mayer, L. 2015, MNRAS, 446, 1957
Förster Schreiber, N. M., Genzel, R., Newman, S. F., et al. 2014, ApJ, 787, 38
Freitag, M., Gürkan, M. A., & Rasio, F. A. 2006, MNRAS, 368, 141
Galli, D., & Palla, F. 1998, A&A, 335, 403
Gammie, C. F. 2001, ApJ, 553, 174
Girichidis, P., Federrath, C., Banerjee, R. & Klessen, R.S., 2011, MNRAS, 413, 2741
Glebbeek, E., Gaburov, E., de Mink, S. E., Pols, O. R., & Portegies Zwart, S. F. 2009, A&A, 497, 255
Gürkan, M. A., Freitag, M., & Rasio, F. A. 2004, ApJ, 604, 632
Hayfield, T., Mayer, L., Wadsley, J., & Boley, A. C. 2011, MNRAS, 417, 1839
Helled, R., Bodenheimer, P., Podolak, M., et al. 2013, arXiv:1311.1142
Hoyle, F. & Fowler, W.A., 1963. MNRAS, 125, 169
Hernquist, L. 1990, ApJ, 356, 359
Hernquist, L. 1993, ApJS, 86, 389
Hollenbach, D., & McKee, C. F. 1979, ApJS, 41, 555
Hopkins, P. F. 2013, MNRAS, 428, 2840
Hosokawa, T., Omukai, K., & Yorke, H. W. 2012, ApJ, 756, 93
Hosokawa, T., Yorke, H.W., Inayoshi. K., Omukai, K., & Yoshida, N., 2013, ApJ, 778, 178
Johnson, J.L., & Bromm, V., 2007, MNRAS, 374, 1557
Kaufmann, T., Mayer, L., Wadsley, J., Stadel, J., & Moore, B. 2007, MNRAS, 375, 53
Kazantzidis, S., Mayer, L., Colpi, M., et al. 2005, ApJ, 623, L67
Keller, B. W., Wadsley, J., Benincasa, S. M., & Couchman, H. M. P. 2014, MNRAS, 442, 3013
Khochfar, S., & Burkert, A. 2006, A&A, 445, 403
Kritsuk, A. G., Norman, M. L., & Wagner, R. 2011, ApJ, 727, L20
Krumholz, M. R., Matzner, C. D., & McKee, C. F. 2006, ApJ, 653, 361
Latif, M. A., Schleicher, D. R. G., & Spaans, M. 2012, A&A, 540, A101
Latif, M. A., Schleicher, D. R. G., Schmidt, W., & Niemeyer, J. 2013, MNRAS, 433, 1607
Lodato, G., & Natarajan, P. 2006, MNRAS, 371, 1813
Lupi, A., Colpi, M., Devecchi, B., Galanti, G., & Volonteri, M. 2014, MNRAS, 442, 3616
Madau, P., & Rees, M. J. 2001, ApJ, 551, L27
Madau, P., Haardt, F., & Dotti, M. 2014, ApJ, 784, L38
Mayer, L., & Wadsley, J. 2004, MNRAS, 347, 277
Mayer, L., Wadsley, J., Quinn, T., & Stadel, J. 2005, MNRAS, 363, 641
Mayer, L., Kazantzidis, S., Madau. P., Colpi, M., Quinn, T. & Wadsley, J. 2007, Science, 316, 1874
Mayer, L., Kazantzidis, S., Escala, A., & Callegari, S. 2010, Nature, 466, 1082
Mayer, L. 2013, Classical and Quantum Gravity, 30, 244008
Meru, F., & Bate, M. R. 2011a, MNRAS, 410, 559
Meru, F., & Bate, M. R. 2011b, MNRAS, 411, L1
Milosavljević, M., Bromm, V., Couch, S. M., & Oh, S. P. 2009, ApJ, 698, 766
Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319
Moody, C. E., Guo, Y., Mandelker, N., et al. 2014, MNRAS, 444, 1389
Montero, P., Janka, H.T. & Müller, E. 2012, ApJ, 749, 37
Mortlock, D. J., et al., 2011, Nature, 464, 616
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Noguchi, M. 1998, Nature, 392, 253
Padoan, P., Nordlund, A., & Jones, B.J.T., 1997, MNRAS, 288, 145
Pelupessy, F. I., Di Matteo, T., & Ciardi, B. 2007, ApJ, 665, 107
Pickett, B. K., Durisen, R. H., & Davis, G. A. 1996, ApJ, 458, 714
Portegies Zwart, S. F., Baumgardt, H., Hut, P., Makino, J., & McMillan, S. L. W. 2004, Nature, 428, 724
Regan, J. A., & Haehnelt, M. G. 2009a, MNRAS, 393, 858

Regan, J. A., & Haehnelt, M. G. 2009b, MNRAS, 396, 343
Reisswig, C., Ott, C. D., Abdikamalov, E., et al. 2013, Physical Review Letters, 111, 151101
Rice, W. K. M., Armitage, P. J., Bate, M. R., & Bonnell, I. A. 2003, MNRAS, 339, 1025
Ritchie, B. W., & Thomas, P. A. 2001, MNRAS, 323, 743
Roškar, R., Fiacconi, D., Mayer, L., et al. 2015, MNRAS, 449, 494
Saijo, M., & Hawke, I. 2009, Phys. Rev. D, 80, 064001
Saitoh, T.R. & Makino, J. 2013, ApJ, 768, 44
Scalo, J., Vázquez-Semadeni, E., Chappell, D., & Passot, T. 1998, ApJ, 504, 835
Schleicher, D. R. G., Palla, F., Ferrara, A., Galli, D., & Latif, M. 2013, A&A, 558, A59
Shen, S., Wadsley, J., & Stinson, G. 2010, MNRAS, 407, 1581
Shibata, M., & Shapiro, S. L. 2002, ApJ, 572, L39
Spaans, M., & Silk, J. 2000, ApJ, 538, 115
Stahler, S. W., & Palla, F. 2005, The Formation of Stars, by Steven W. Stahler, Francesco Palla, pp. 865. ISBN 3-527-40559-3. Wiley-VCH , January 2005.
Stinson, G., Seth, A., Katz, N., et al. 2006, MNRAS, 373, 1074
Tacconi, L. J., Genzel, R., Neri, R., Cox, P., Cooper, M. C., Shapiro, K., Bolatto, A., Bouché, N., Bournaud, F., Burkert, A., Combes, F., Comerford, J., Davis, M., Förster Schreiber, N. M., Garcia-Burillo, S., Gracia-Carpio, J., Lutz, D., Naab, T., Omont, A., Shapley, A., Sternberg, A., & Weiner, B. 2010, Nature, 463, 781
Tanaka, T., & Haiman, Z. 2009, ApJ, 696, 1798
Treister, E., Schawinski, K., Volonteri, M., & Natarajan, P. 2013, ApJ, 778, 130
Van Wassenhove, S., Capelo, P. R., Volonteri, M., et al. 2014, MNRAS, 439, 474
Vink, J. S. 2008, New Astron. Rev., 52, 419
Volonteri, M., Haardt, F., & Madau, P. 2003, ApJ, 582, 559
Volonteri, M., & Rees, M. J. 2006, ApJ, 650, 669
Volonteri, M., & Begelman, M. C. 2010, MNRAS, 409, 1022
Walter, F., et al., 2004, ApJ, 615, L17
Wadsley, J. W., Stadel, J., & Quinn, T. 2004, New A, 9, 137
Wise, J. H., Turk, M. J., & Abel, T. 2008, ApJ, 682, 745
Willott, C., et al., 2010, AJ, 139, 906

Mon. Not. R. Astron. Soc. **000**, 1–13 (2015) Printed April 11, 2016 (MN LaTeX style file v2.2)

Bar-formation as driver of gas inflows in isolated disc galaxies

R. Fanali[1], M. Dotti[1,2], D. Fiacconi[3] & F. Haardt[4,2]
[1] *Università degli Studi di Milano Bicocca, I–20126 Milano, Italy*
[2] *INFN, Sezione di Milano–Bicocca, Piazza della Scienza 3, I–20126 Milano, Italy*
[3] *Institute for Computational Science, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[4] *DiSAT, Università dell'Insubria, via Valleggio 11, I–22100 Como, Italy*

April 11, 2016

ABSTRACT

Stellar bars are a common feature in massive disc galaxies. On a theoretical ground, the response of gas to a bar is generally thought to cause nuclear starbursts and, possibly, AGN activity once the perturbed gas reaches the central super-massive black hole. By means of high resolution numerical simulations we detail the purely dynamical effects that a forming bar exerts on the gas of an isolated disc galaxy. The galaxy is initially unstable to the formation of non-axisymmetric structures, and within ~ 1 Gyr it develops spiral arms that eventually evolve into a central stellar bar on kpc scale. A first major episode of gas inflow occurs during the formation of the spiral arms while at later times, when the stellar bar is establishing, a low density region is carved between the bar co-rotational and inner Lindblad resonance radii. The development of such "dead zone" inhibits further massive gas inflows. Indeed, the gas inflow reaches its maximum during the relatively fast bar formation phase and not, as often assumed, when the bar is fully formed. We conclude that the low efficiency of long-lived, evolved bars in driving gas toward galactic nuclei is the reason why observational studies have failed to establish an indisputable link between bars and AGNs. On the other hand, the high efficiency in driving strong gas inflows of the intrinsically transient process of bar formation suggests that the importance of bars as drivers of AGN activity in disc galaxies has been overlooked so far. We finally prove that our conclusions are robust against different numerical implementations of the hydrodynamics routinely used in galaxy evolution studies.

Key words: galaxies: bulges — galaxies: nuclei — methods: numerical

1 INTRODUCTION

The fraction of disc galaxies showing a well developed stellar bar in the local Universe is substantial, up to $\gtrsim 30\%$ for massive ($M_* \gtrsim 10^{9.5}\mathrm{M}_\odot$) systems (Laurikainen, Salo & Buta 2004; Nair & Abraham 2010; Lee et al. 2012a; Gavazzi et al. 2015). The effectiveness of bars in modifying the dynamics of gas has been recognized since decades (e.g. Sanders & Huntley 1976; Roberts, Huntley & van Albada 1979; Athanassoula 1992). In particular, gas within the bar corotational radius (R_C, i.e. the radius at which the angular velocity in the disc plane $\Omega(R)$ equals the bar pattern precession speed Ω_b) is driven toward the centre of the galaxy because of the interaction with the bar itself. Early theoretical studies suggested that such inflows could be responsible for nuclear starbursts and, if the gas is able to reach the very central regions of the galaxy, AGN activity (e.g. Shlosman, Frank & Begelman 1989; Berentzen et al. 1998).

From the observational point of view the connection between bars and enhanced nuclear star formation has been extensively proved (e.g. Ho, Filippenko & Sargent 1997; Martinet & Friedli 1997; Hunt & Malkan 1999; Laurikainen, Salo & Buta 2004; Jogee, Scoville & Kenney 2005). The link between bars and AGN seems less clear: while barred galaxies host AGNs more frequently that their non-barred analogous (making bars a good candidate for the triggering of nuclear activity, e.g. Laurikainen, Salo & Buta 2004; Oh, Oh & Yi 2012), it is still matter of debate whether the presence of bars is one of the main drivers of AGNs (as suggested by, e.g. Knapen, Shlosman & Peletier 2000; Laine et al. 2002; Alonso, Coldwell & Lambas 2013) or not (see e.g. Ho, Filippenko & Sargent 1997; Mulchaey & Regan 1997; Hunt & Malkan 1999; Lee et al. 2012b; Cisternas et al. 2013; Cheung et al. 2015).

In order to have a comprehensive understanding of the gas dynamics in barred galaxies many numerical studies have been put forward, including, for example both 2- or 3-D simulations, and different schemes for the gas hydrodynamics (smoothed particle hydrodynamics, SPH, vs grid codes). We consider particularly meaningful to divide the different efforts in three main classes:

(i) **Isolated galaxies with analytical bars** (e.g. Athanassoula 1992; Regan & Teuben 2004; Kim, Seo & Kim 2012). In this class of simulations (often restricted to a 2-D geometry) bars are represented by analytical potentials that do not evolve in time (but for their rigid body rotation). These simulations, although quite idealized, allows for extremely high resolutions and precise evolution of the gas dynamics.

(ii) **Fully evolving isolated galaxies** (e.g. Berentzen et al. 1998, 2007; Villa-Vargas, Shlosman & Heller 2010; Cole et al. 2014), where bars are modeled (as the rest of the galaxy) as evolving structures, that can change their extents, rotational patterns, etc.

(iii) **Cosmological simulations** (e.g. Romano-Díaz et al. 2008; Scannapieco & Athanassoula 2012; Kraljic, Bournaud & Martig 2012; Goz et al. 2014; Fiacconi, Feldmann & Mayer 2015). In these simulations the galaxies form from cosmological perturbations, and are free to acquire mass and angular momentum through large scale gas inflows and galaxy mergers. In this approach the initial conditions are not arbitrary, but, because of the large boxes simulated (even in zoom-in runs), the spatial and mass resolution is usually significantly coarser than in isolated simulations.

Simulations of the first kind have confirmed the analytical prediction that, in many galactic potentials, bar-driven gas inflows fail to reach the very centre of the galaxy. The gas shocks around the outermost inner Lindblad resonance (ILR) radius ($R_{\rm ILR}$) of the bar, defined by the equality $\Omega(R) - \kappa(R)/2 = \Omega_b$ where κ is the epicyclic frequency, i.e. the frequency of small radial oscillations. At $R_{\rm ILR}$ the gas shocks, forming nuclear rings that are often observed as star forming regions in barred galaxies (e.g. Kormendy 2013, and references therein). Simulations that fully evolve the bar potential do show similar results as soon as they reach a quasi-steady state, i.e. after the bar growth transient[1]. If the gas inflows accumulates enough mass at $\sim R_{\rm ILR}$ the central region can dynamically decouple, possibly forming nested non-axisymmetric structures (e.g. nuclear bars). These structures can eventually bring the gas closer and closer to the galactic centre in a cascade-like fashion (Shlosman, Frank & Begelman 1989).

In this paper we propose a new set of fully evolving isolated galaxies runs. We start with an unbarred galactic disc composed of stars and gas, embedded in an evolving dark matter halo. We check the dependences of the gas dynamics on different numerical implementations, varying the magnitude of an artificial viscosity (if present) and the numerical resolution (see section 2.2 for a full description of the different runs). We run our simulations without implementing any gas radiative cooling, star formation and stellar feedback prescriptions (usually referred to as sub-grid physics), in order to perform a clean test of the basic numerical method used, and to highlight the physical and purely dynamical effect of the forming substructures (stellar spirals and bar) onto the gas.

As will be detailed we find that the flux of gas reaching the most central regions of the galaxy peaks during the bar formation phase, and not when the bar is fully established, independently of the exact numerical implementation. We describe in details the set-up of our initial conditions and the features of the simulation suite in Section 2. We present our main findings in Section 3, and we finally discuss them and derive our conclusions in Section 4, highlighting the relevance of our work for the interpretation of observations and also commenting on the possible shortcomings.

2 NUMERICAL METHODS

2.1 Initial conditions

We simulate the isolated disc galaxy model Lmd2c12 described by Mayer & Wadsley (2004), in order to reproduce an initially bulgeless, bar-unstable disc galaxy. The galaxy model is made of three different components: a dark matter halo, a stellar and a gaseous disc.

The dark matter halo follows the Navarro, Frenk & White (NFW 1996, 1997) density profile:

$$\rho_{\rm h}(r) = \frac{\rho_{\rm crit}\ \delta_{\rm c}}{(r/r_{\rm s})(1+r/r_{\rm s})^2}, \qquad (1)$$

where $r_{\rm s}$ is the scale radius of the halo, $\rho_{\rm crit}$ is the critical density of the Universe today[2], and:

$$\delta_{\rm c} = \frac{200}{3}\frac{c^3}{\log(1+c) - c/(1+c)}, \qquad (2)$$

depends only on the concentration parameter $c \equiv r_{200}/r_{\rm s}$. r_{200} is the radius that encompasses an average density $\langle\rho\rangle = 200\ \rho_{\rm c}$ and defines the outer radius of the dark matter halo. The mass of the halo is therefore $M_{200} = 200\ \rho_{\rm c}\ (4\pi/3)\ r^3_{\rm vir}$. We adopt $c = 12$ and a scale velocity $v_{200} = \sqrt{GM_{200}/r_{200}} = 75$ km s^{-1}, which corresponds[3] to $M_{200} = 1.4 \times 10^{11}$ M$_{\odot}$, $r_{200} = 110$ kpc and $r_{\rm s} = 9.2$ kpc.

Both stellar and gaseous discs are modeled as a radial exponential disc with a vertical structure modelled by isothermal sheets (Hernquist 1993):

$$\rho_\star(R,z) = \frac{M_\star}{4\pi R_\star^2 z_\star}\exp(-R/R_\star)\ \cosh^{-2}\left(\frac{z}{z_\star}\right), \qquad (3)$$

where $R_\star = 3$ kpc is the radial scale length and $z_\star = 0.3$ kpc is the vertical scale height. The stellar disc has a total stellar mass $M_\star = 1.4\times10^{10}$ M$_{\odot}$ and extends up to $10R_\star$. The gas component has a mass $M_{\rm gas} = 0.05M_\star = 7\times10^8$ M$_{\odot}$ and its density profile is characterized by the same parameters $R_\star$ and $z_\star$. The gas has a uniform temperature $T_0 = 10000$ K

---

[1] Although promising, the coarse resolution of cosmological runs makes hard to fully resolve the nuclear region where the ILR is expected to occur.

[2] We assume $H_0 = 71$ km s^{-1} Mpc^{-1}, compatible with the *Wilkinson Microwave Anisotropy Probe* 7/9 years cosmology (Komatsu et al. 2011; Hinshaw et al. 2013)

[3] These numbers are slightly different from those reported by Mayer & Wadsley (2004) because of the different cosmology assumed. However, this does not affect the evolution of the galaxy model.

and we assume that it is composed of a mixture of ionized hydrogen and helium with a mean molecular weight $\mu \simeq 0.59$. All the parameters are chosen in agreement with the galaxy-halo scalings predicted by the Λ-CDM model (e.g. Mo, Mao & White 1998).

We build the initial conditions using the code GINCO[4]. GINCO initializes quasi-equilibrium galaxy models following Hernquist (1993) and Springel, Di Matteo & Hernquist (2005). The models can be made of four arbitrary components: a NFW dark matter halo, an exponential stellar and gaseous disc, and a spherical bulge with the profile proposed by Hernquist (1990). The polar/spherical coordinates of the particles that belong to each component are randomly sampled using the density profiles as probability distribution functions. Then, polar/spherical angles are randomly drawn from isotropic distributions and they are used to determine the Cartesian coordinates of the particle positions.

The velocities are sampled from local Gaussian approximations of the true distribution function (Hernquist 1993). The position-dependent parameters of the Gaussians are computed solving the steady-state Jeans equations with some closure assumptions on the velocity dispersion tensor (see e.g. Binney & Tremaine 2008). For spherical components (i. e. the dark matter halo, since the simulated system is bulgeless), we assume that the velocity dipersion tensor is isotropic (i.e. of the form $\sigma^2(r)\,\mathbb{I}$, where $\mathbb{I}$ is the identity matrix), with the 1D velocity dispersion given by:

$$\sigma^2(r) = \frac{1}{\rho(r)} \int_r^{+\infty} \rho(x) \frac{\mathrm{d}\Phi_{\mathrm{tot}}}{\mathrm{d}r}(x)\ \mathrm{d}x, \tag{4}$$

where $\rho(r)$ is the density profile of the considered component and Φ_{tot} is the total gravitational potential. The spherical components have no net rotation. The potential of the halo is an analytic function; instead, the potential of the disc is computed as a first-order vertical perturbation of the potential of a razor-thin exponential disc, namely $\Phi_{\mathrm{d}}(R,z) \simeq \Phi_0(R) + \Phi_1(R,z)$. The razor-thin disc has the potential:

$$\Phi_0(R) = -\frac{GM_\star}{R_\star}\ y\ \left[I_0(y)K_1(y) - I_1(y)K_0(y)\right], \tag{5}$$

where $y = R/(2R_\star)$ and I_i and K_i are modified Bessel functions; the first-order vertical perturbation is (e.g. Binney & Tremaine 2008):

$$\begin{aligned} \Phi_1(R,z) &\equiv 4\pi G \int_0^z \mathrm{d}z' \int_0^{z'} \mathrm{d}z'' \rho_\star(R,z'') \\ &= 4\pi G\ \rho(R,0)\ z_\star^2\ \log\left[\cosh\left(\frac{z}{z_\star}\right)\right]. \end{aligned} \tag{6}$$

We use this strategy to maintain all the evaluations of the potentials and of their derivatives analytic; this makes the code faster and reduces the required memory. Once we compute $\sigma^2(r)$, we can sample the magnitude of the velocity of each particle in a spherical component from a Maxwellian distribution with variance $\sigma^2(r)$. Finally, we randomly draw the spherical angles (θ, ϕ) as above to ensure isotropy and we assign the Cartesian components of the velocity.

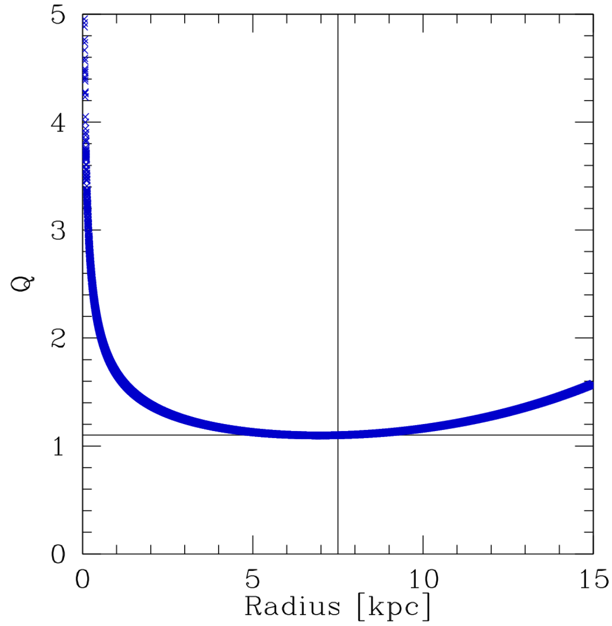

Figure 1. Radial profile for the Toomre parameter of the initial stellar disc.

Both the stellar and the gaseous disc velocity structure is characterized by a velocity dispersion tensor of the form $\mathrm{diag}(\sigma_r^2, \sigma_\phi^2, \sigma_z^2)$. The vertical velocity dispersion is (Hernquist 1993; Springel, Di Matteo & Hernquist 2005):

$$\begin{aligned} \sigma_z^2(R,z) &= \frac{1}{\rho_\star(R,z)} \int_z^{+\infty} \rho_\star(R,z') \frac{\partial \Phi_{\mathrm{tot}}}{\partial z}(R,z')\ \mathrm{d}z' \\ &\approx \frac{GM_\star z_\star}{2R_\star^2} \exp\left(-\frac{R}{R_\star}\right), \end{aligned} \tag{7}$$

where the last approximation holds when the disc is geometrically thin and the vertical gradient of the potential around $z \simeq 0$ is dominated by the disc. The radial component is chosen to be $\sigma_r^2 \propto \sigma_z^2$, with the normalization enforcing a minimum value of the Toomre parameter $\mathcal{Q} \simeq 1.1$ at $r \simeq 2.5R_\star$ (Mayer & Wadsley 2004). The whole profile of $\mathcal{Q}$ for our initial conditions is shown in Fig. 1. The azimuthal component is set using the epicyclic approximation, $\sigma_\phi^2 = \sigma_r^2\ \kappa^2/(4\Omega^2)$. Unlike the dark matter halo, the disc has a net rotation, i.e. an average azimuthal velocity $\langle v_\phi \rangle$ given by (Hernquist 1993; Springel, Di Matteo & Hernquist 2005):

$$\langle v_\phi \rangle^2 = V_{\mathrm{c}}^2 + \sigma_r^2 \left(1 - \frac{\kappa^2}{4\Omega^2} - \frac{2R}{R_\star}\right), \tag{8}$$

where V_{c} is the circular velocity in Φ_{tot}. Finally, we sample the (v_r, v_ϕ, v_z) components of the velocity of each disc (both star and gas) particles from gaussian distributions with mean $(0, \langle v_\phi \rangle, 0)$ and standard deviations $(\sigma_r, \sigma_\phi, \sigma_z)$, respectively, and we finally transform then into the Cartesian components.

We checked the stability of our initial conditions studying the evolution of the stellar surface density profile as a function of time (left panel of Fig. 2) during the first Gyr, i.e. before the development of strong non-axisymmetric perturbation (see below). After a short transient phase due to the non-exact equilibrium of the initial conditions (highlighted

---

[4] GINCO (Galaxy INitial COnditions, `http://www.ics.uzh.ch/~fiacconi/software.html`) was written by Davide Fiacconi.

be the yellow line in figure) the system re-adjust on a profile similar to the initial one, with the surface density at $t = 1$ Gyr (red line) differing by 20% at most with respect to the initial conditions (within the disc scalelength). Similar conclusions about the stability of the stellar disk can be drawn from the evolution of its Lagrangian radii (right panel of Fig. 2).

2.2 Simulation suite

We run a suite of numerical simulations of the reference model described in the previous Section in order to explore the effects of resolution, numerical implementation and parametrization of the artificial viscosity (when present). In Table 1 we summarize the sample of 3D runs presented in this work. We build two realizations of our initial conditions at two resolutions:

(i) low resolution (LR): the halo is sampled with 10^6 particles with mass $m_{\rm h} = 1.4 \times 10^5$ M$_\odot$, while the stellar and gaseous discs are sample with 9.5×10^5 and 5×10^4 particles, respectively, with mass $m_\star = m_{\rm gas} \simeq 1.5 \times 10^4$ M$_\odot$. The gravitational softenings (setting the spatial/force resolution of the gravitational interaction) for dark matter and baryonic particles (equal for stars and gas particles) are 65 pc and 20 pc, respectively;

(ii) high resolution (HR): the halo is sampled with 8×10^6 particles with mass $m_{\rm h} = 1.6 \times 10^4$ M$_\odot$, while the stellar and gaseous discs are sample with 7.6×10^6 and 4×10^5 particles, respectively, with mass $m_\star = m_{\rm gas} \simeq 1.7 \times 10^3$ M$_\odot$. The gravitational softenings for dark matter and baryonic particles are 30 pc and 7 pc, respectively.

We ensure that the particles in the disc (star and gas) have all the same mass, preventing any spurious relaxation/mass segregation. All the simulations assume an isothermal equation of state to simply model an effective atomic radiative cooling keeping the ISM in the disc plane at an almost constant temperature $\lesssim 10^4$ K. Metal line and molecular cooling would reduce the gas temperature further, allowing for dense clumps to form and to trigger star-formation. Feedback from stars would then re-heat the gas, resulting in the formation of a multi-phase medium(e.g. Wada 2001; Wada & Norman 2001). Because of the lack of cooling and star-formation physics, we keep an high temperature to prevent the sudden fragmentation of the gaseous disc.

We test the robustness of our results against two different implementations of the hydrodynamics. Most of the simulations are performed with the Tree/Smoothed Particle Hydrodynamics (SPH) code GADGET2 (Springel 2005), which uses an oct-tree structure to speed up the gravity calculations (Barnes & Hut 1986) and threats the hydrodynamics with the density-entropy SPH proposed by Springel & Hernquist (2002). The SPH formalism requires the introduction of an artificial viscosity in order to capture shocks correctly (e.g. Monaghan 1992; Balsara 1995; Monaghan 1997). Therefore, we explore the effect of different choices of the value of the artificial viscosity parameter α[5]. Finally, we also compare the results from SPH simulations with a run that uses the newly developed code GIZMO (Hopkins 2014). GIZMO is a meshfree code that captures advantages from both SPH and grid codes: it preserves the Lagrangian structure of SPH codes, but at the same time solved directly the Euler equations among different regions of the computational domain without requiring the implementation of any artificial viscosity. We used it in its finite-mass variant, in which there is not mass flux among the regions belonging to different particles, keeping the mass of each gas particle fixed.

Table 1. Summary of simulations characteristics. Columns: (1) name of the simulation, (2-3) resolution, (4) code used, (5) artificial viscosity α.

Name	Barion particle softening (pc)	DM particle softening (pc)	Code	α
LR	20	65	GADGET2	0.8
LRV16	20	65	GADGET2	1.6
LRV04	20	65	GADGET2	0.4
LRGiz	20	65	GIZMO	-
HR	7	30	GADGET2	0.8

3 RESULTS

3.1 Low resolution simulations

Figure 3 shows the distribution of star and gas observed in the LR run at three different times, t =1, 4 and 7 Gyr in the left, central and right panels respectively. The stellar surface density is shown in the upper and middle panels (edge-on and face-on views, respectively), while the face-on view of the gas surface density is shown in the lower panels. During the first 2 Gyr the bar-unstable system evolves from a axisymmetric configuration to a barred disc, passing through the formation of transient multi-arm spirals. In particular, a three arm spiral structure is observable in the stellar density distribution at t =1 Gyr in the left-middle panel of figure 3. From 2 Gyr on the disc shows a clear bar structure (with a size of about 8 kpc) in its central region. From the bar-formation time ($t \approx 2$ Gyr) on, the bar tends to slow-down, as shown in the upper panel of figure 4. At $t \lesssim 3$ Gyr the bar makes almost 2.8 full precessions per Gyr, while the frequency decreases down to $\lesssim 2.3$ precessions per Gyr at $t \approx 7$ Gyr. The bar slow-down, already extensively discussed in literature (e.g. Sellwood 1981; Combes & Sanders 1981; Halle et al. 2015), results in a $R_{\rm ILR}$ growing in time, from ~ 1 kpc up to ~ 1.4 kpc at the end of the run, as observable in the lower panel of figure 4. The bar forms thin, and buckles in its centre as the time goes by, as observable in the edge on view of the stellar disc at t =4 and 7 Gyr. At the end of the simulation a boxy-peanut bulge like structure is observable within the central few kpc of the disc.

The dynamics of the subdominant gas component is dominated by the underlying stellar dynamics. During the first 2 Gyr the gas distribution resembles the stellar one, with clear spiral arms (almost co-spatial with the stellar ones) observable (see the example in the left lower panel of figure 3 at t =1 Gyr). After the formation of the stellar bar, the gas within the bar corotational radius ($R_{\rm C} \approx$ 4-5 kpc depending on the age of the bar, as will be discussed below)

---

[5] The β parameter in the Monaghan-Balsara formulation is equal to 2 α in all our runs.

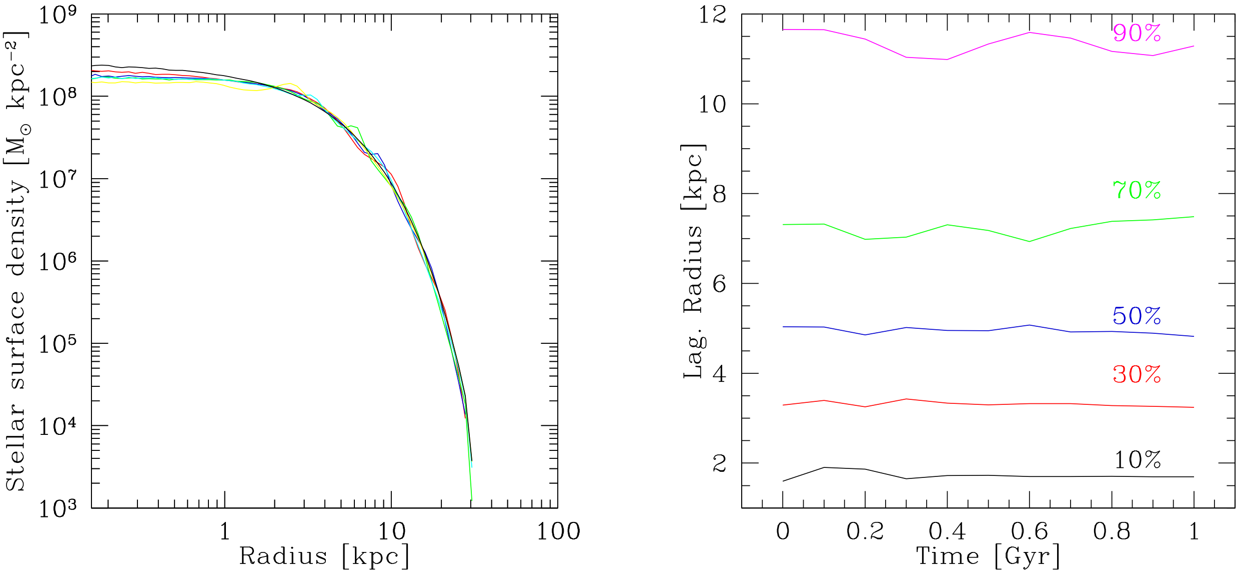

Figure 2. Left panel: surface density profile of the stellar component in the first Gyr. Black, yellow, cyan, green, blue and red curves correspond to t =0, 0.2, 0.4, 0.6, 0.8 and 1 Gyr, respectively. Right panel: lagrangian radius at different stellar mass fraction in the first Gyr. Black, red, blue, green and magenta lines represent 10%, 30%, 50%, 70%, 90% of stellar mass, respectively.

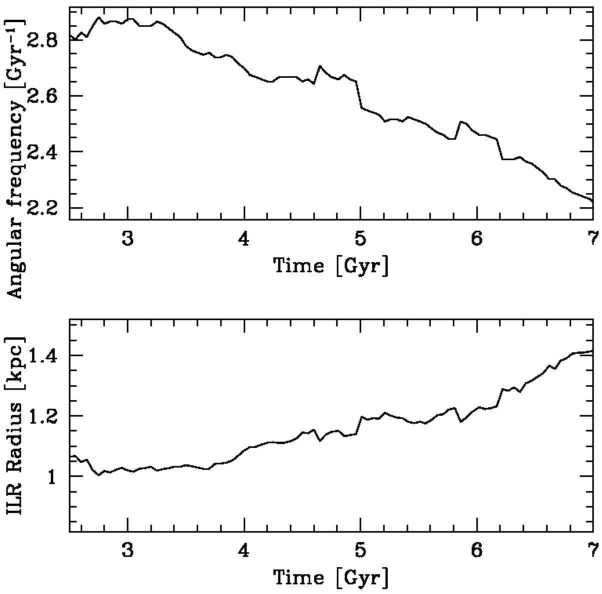

Figure 4. Angular frequency of the bar (upper panel) and radius corresponding to the inner Lindblad resonance from t=2.5 to t=7 Gyr (lower panel).

is driven toward the galaxy centre, and forms a dense knot of gas clearly observable in the central and right panels in the bottom row of figure 3. The torquing effect of the spiral arms before and the stellar bar afterwards sweeps the almost totality of the gas between $R_{\rm C}$ and the central dense knot. A small amount of low dense gas is still observable in this "dead region", in particular in the form of two inflowing streams connecting the outer galactic disc with the central dense knot, often observed in simulations as well as in real galaxies (e.g. Regan, Sheth & Vogel 1999).

The left panel of figure 5 quantifies the effect that the bar formation process has onto the gas. The surface density of the gas in the dead zone decreases by up to ~ 1.5 orders of magnitude at $t \gtrsim 3$ Gyr (blue, yellow and red lines) with respect to the initial conditions (black line). The shaded areas in figure trace the evolution of $R_{\rm C}$ (green) and its outermost inner Lindblad resonance radius ($R_{\rm ILR}$ pink), from when a clear bar structure is observable and its angular frequency is measurable ($t \approx 2$ Gyr) to the end of the simulation. The gas within $R_{\rm C}$ is dragged toward scales of the order of $R_{\rm ILR}$, fueling the formation of the central knot of gas on sub-kpc scales (in agreement with a wealth of previous studies, e.g. Sanders & Huntley 1976; Shlosman, Frank & Begelman 1989; Athanassoula 1992; Berentzen et al. 1998; Regan & Teuben 2004; Kim, Seo & Kim 2012; Cole et al. 2014), where the surface density increases by up to almost 2 orders of magnitude.

A clear although less obvious result of the LR run consists in the efficiency of the "dead zone" formation. Most of the inflow from $R < R_{\rm C}$ to $R \lesssim R_{\rm ILR}$ happens during the first 2 Gyr, as observable comparing the cyan (t =1 Gyr), green (t =2 Gyr) and blue (t =3 Gyr) lines with the initial conditions and the end result of the simulation in the left panel of figure 5. The fully formed bar does indeed play a role in further decreasing the gas surface density on the dead zone, and most importantly, in preventing new gas to refill the central regions by pushing the gas immediately outside the CR toward the outer Lindblad resonance radius ($R_{\rm OLR}$).[6] However, it is instead the formation of the bar

---

[6] Although harder to be noticed in a log-log plot, the gas surface density decreases in the $R_{\rm C} < R < R_{\rm OLR}$ region, and the material accumulates just outside $R_{\rm OLR}$.

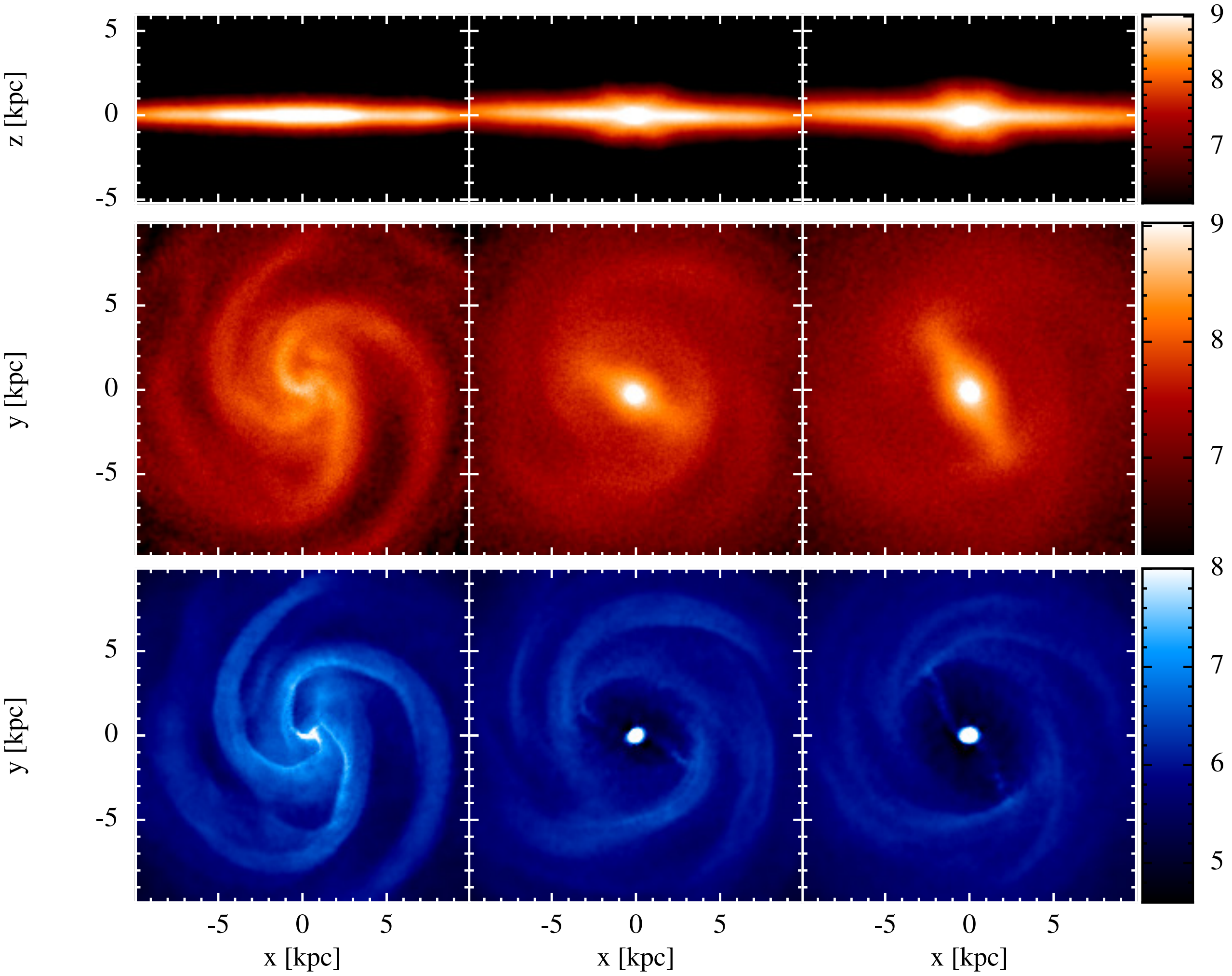

Figure 3. Upper (middle) panels: edge-on (face-on) views of the stellar disc at t =1, 4 and 7 Gyr (left, central and right panel, respectively). The colour gradient maps the stellar surface density (in units of $M_\odot$ kpc^{-2}) on a logarithmic scale. Bottom panels, same as the middle panel for the gas surface density.

which is efficient in driving sustantial gas inflow. The fundamental importance of the torques acting on the gas during the build-up of the bar, before this has been fully developed, is highlighted in the right panel of figure 5, in which we show the gas accretion rate $\dot{M}$ within R_C as a function of time. In particular, the red and blue lines refer to $\dot{M}$ through surfaces at 0.3 and 1 kpc from the centre, respectively. At both scales $\dot{M}$ shows a first prominent peak at $t \approx 1$ Gyr, well before the formation of any significant bar-like structure. A second peak of similar magnitude is observable at $t \approx 2$ Gyr, just after the bar has formed, while the central fueling drops immediately afterward. Although our simple simulation does not include any star formation prescription, such omission has little impact on the evolution of the gas from R_C down to the nuclear knot, since the majority of the inflow happens on a few (up to ≈ 10 close to $R_{\rm ILR}$) orbital timescales. The time evolution of the accretion flows through the two surfaces is quite similar at all times.

3.2 Viscosity test

As recently reviewed by Sellwood (2014), the gas angular momentum transport in SPH simulations could be at least partially affected by the artificial viscosity used. Differently from grid based codes, in which a numerical viscosity is intrinsically related with the discretization of the space domain, in SPH codes the numerical viscosity is explicitly taken into account through a viscosity parameter α. The shear and bulk viscosity in SPH simulations scale linearly with α (e.g. Murray 1996; Lodato & Price 2010).

In this section we test the effect of the artificial viscosity through the comparison of run LR with three different simulations. Two of these, LRV04 and LRV16, are exact copies

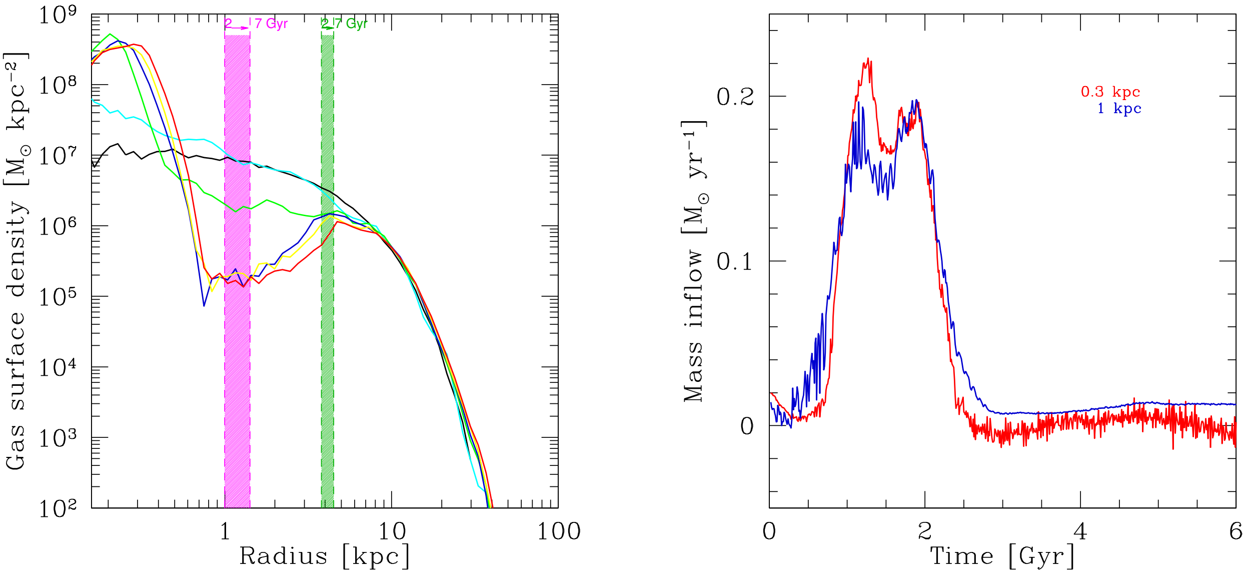

Figure 5. Left panel: surface density profile of the gaseous component. Solid black, cyan, green, blue, yellow and red curves correspond to t =0,1,2,3,5 and 7 Gyr, respectively. The shaded magenta and green areas mark the region span by $R_{\rm ILR}$ and $R_{\rm C}$, respectively. Right panel: gaseous mass inflow as a function of time. Red and blue lines represent the accretion rate computed at 0.3 and 1 kpc from the center respectively.

of the LR run, but for the value of the α parameter, that is half and double of the $\alpha = 0.8$ value used in LR. The third simulation (LRGiz) has been run using the GIZMO code (Hopkins 2014), that solves the evolution of the gas on an unstructured grid and does not require any explicit artificial viscosity term.

The results of the test are shown in figure 6 and figure 7. Figure 6 shows the comparison between the surface density profiles of the four runs (LR in red, LRV16 in green, LRV04 in blue and LRGiz in cyan) at four diffent times, t =1 Gyr (upper left panel), 3 Gyr (upper right panel), 5 Gyr (lower left panel) and 7 Gyr (lower right panel). Similarly, figure 7 shows the face-on projection of the gas surface density map for the four runs at t = 1 Gyr, to allow for a comparison of the non-axisymmetric structures forming. The comparison between the three SPH runs shows that the exact value of the viscosity parameter α plays a little role in the gas dynamics. The removal of gas from the forming dead zone and the formation of a dense central gas knot are completely dominated by the gravitational torques due to the formation of non-axisymmetric structures. The LRGiz run shows some very minor differences too.

3.3 High resolution simulations

3.3.1 Low vs high resolution run comparison

As a final test, we ran an increased resolution version of the LR simulation (HR), as discussed in section 2. Because of the higher spatial resolution and of the isothermal equation of state implemented, the gas in the HR run forms extremely dense and compact clouds in the galaxy nucleus, slowing down the simulation enormously after the first episode of major gas inflow. For this reason we have run the HR simulation only up to $t \approx 3$ Gyr, and we limit our analysis to the response of the gas to the initial spiral and bar formation.

Fig. 8 shows the face-on views of the stellar (left panels) and gaseous (right panels) surface density. Similarly to the low resolution simulation, during the first Gyr, the system develops stellar spiral arms (upper left panel) which evolve in a stable bar like structure (lower left panel) at about t=2.5 Gyrs. The gas follows a similar dynamics as observed in the lower resolution runs, following the stellar spiral arms during the first evolutionary phase and being driven toward the centre during and after the bar formation. As in the other runs, in the central region of the galaxy (within the bar extent) a dead zone forms, with very low density gas present in between the outer disc and the nuclear gas knot.

Figure 9 shows a comparison between HR and LR runs. The left panel shows the gas surface density profiles in the two runs for four different times. The biggest difference is observable at $t \approx 1$ Gyr, when in the low resolution run the disc is already significantly perturbed, while in the HR run the gas profile is still quite unperturbed. The gas profile in the high resolution run is more similar to its low resolution analogous at later times, but for a slightly more pronounced dead zone in the HR run due to the better resolved profile of the stellar bar, that results in a more effective action of the bar itself onto the gas.

The main difference observed in the profiles at $t \approx 1$ Gyr is due to the later growth of non-axisymmetric perturbations (first in the form of spiral arms, turning into a central bar) in the higher resolution simulation. This is clearly observable in the accretion rate through the central 0.3 kpc (right panel of figure 9). In run HR the peak of $\dot{M}$ occurs at $t \approx 1.5$ Gyr, about 0.2-0.3 Gyr after the peak observed in the LR run. Again, in the high resolution simulation the $\dot{M}$ peak has a larger intensity (by almost a factor of 2) with respect to the

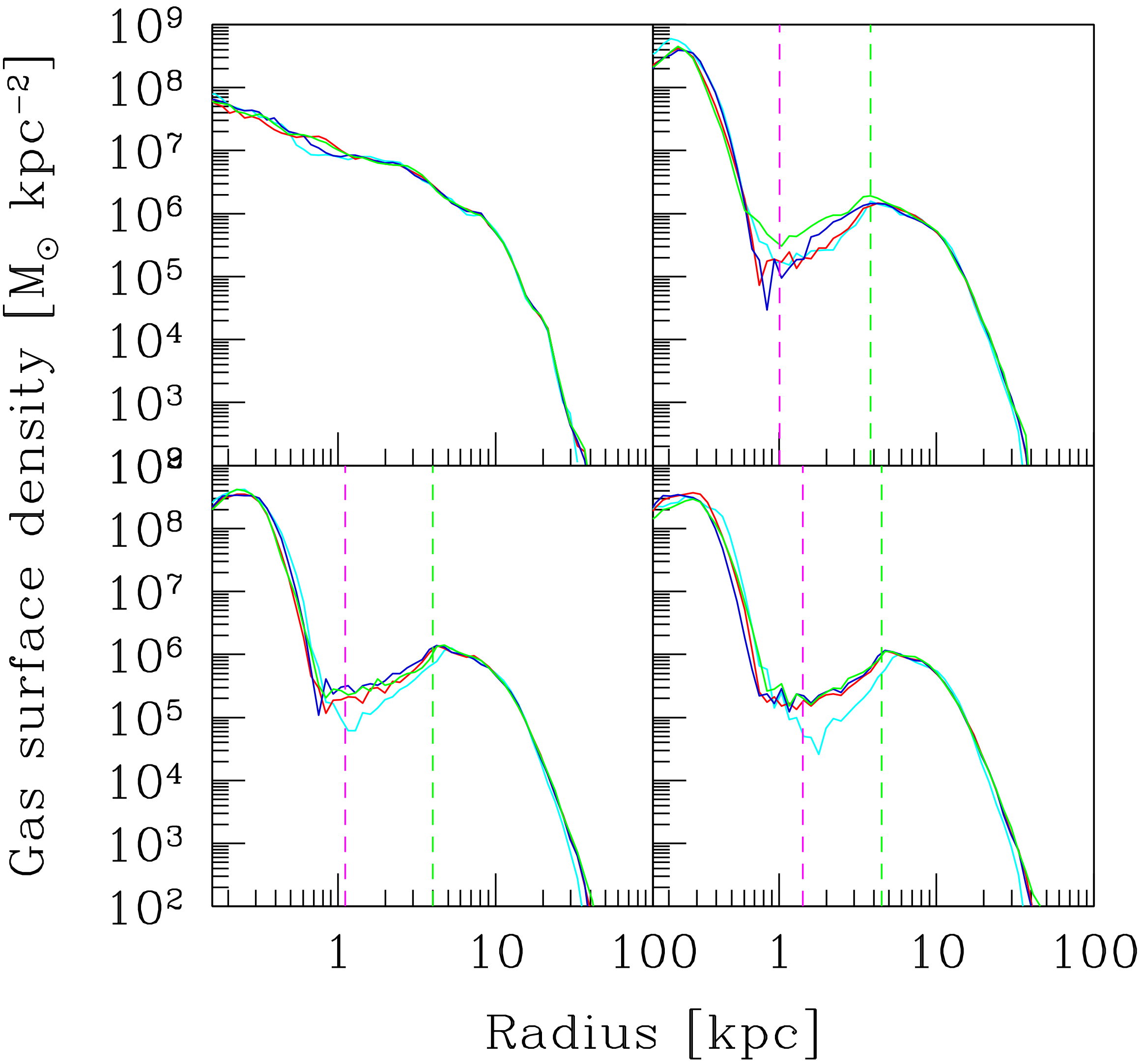

Figure 6. Surface density profile for low resolution isothermal simulations with different viscosity. Upper left, upper right, lower left and lower right panels refer to t=1 Gyr, 3 Gyrs, 5 Gyrs and 7 Gyrs, respectively. The solid red, green and blue curves are obtained from the gas particle distribution for simulation with α = 0.8 (run LR), α = 1.6 (run LRV16) and α = 0.4 (run LRV04). The cyan curve corresponds to the gas particle distribution for simulation using GIZMO (LRGiz). The dashed magenta and green lines mark the positions of $R_{\rm ILR}$ and $R_{\rm C}$ at the different times.

low resolution case, due to the more efficient cleaning of the dead zone during the bar formation process.

The later growth of non-axisymmetric structures in the higher resolution run is probably due to the lower shot noise in the initial conditions: a higher number of particle MonteCarlo sampling results in a lower statistical noise, from which structures can grow (see also the discussion in Sellwood 2014). An extensive and time consuming study aiming at numerical convergence is neither feasible (within the currently available computational facilities) nor useful, as a simulation with a order of magnitudes larger number of particles could result in a degree of symmetry significantly larger than any real disc galaxy observed. The dependence of the $\dot{M}$ peak and of the time at which spirals and bars form on the number of particles used demonstrate that these should not be taken as physical values. Only the gas response pattern is similar in all the runs analyzed, independently of the viscosity prescription adopted, of the numerical resolution achieved, and of the algorithm used to solve the gas dynamics.

3.3.2 Dynamics of the nuclear inflow

The high resolution achieved allows us to resolve sub-kpc scales, and to investigate the detailed causes of the nuclear gas inflow through the search of stellar and gaseous nuclear

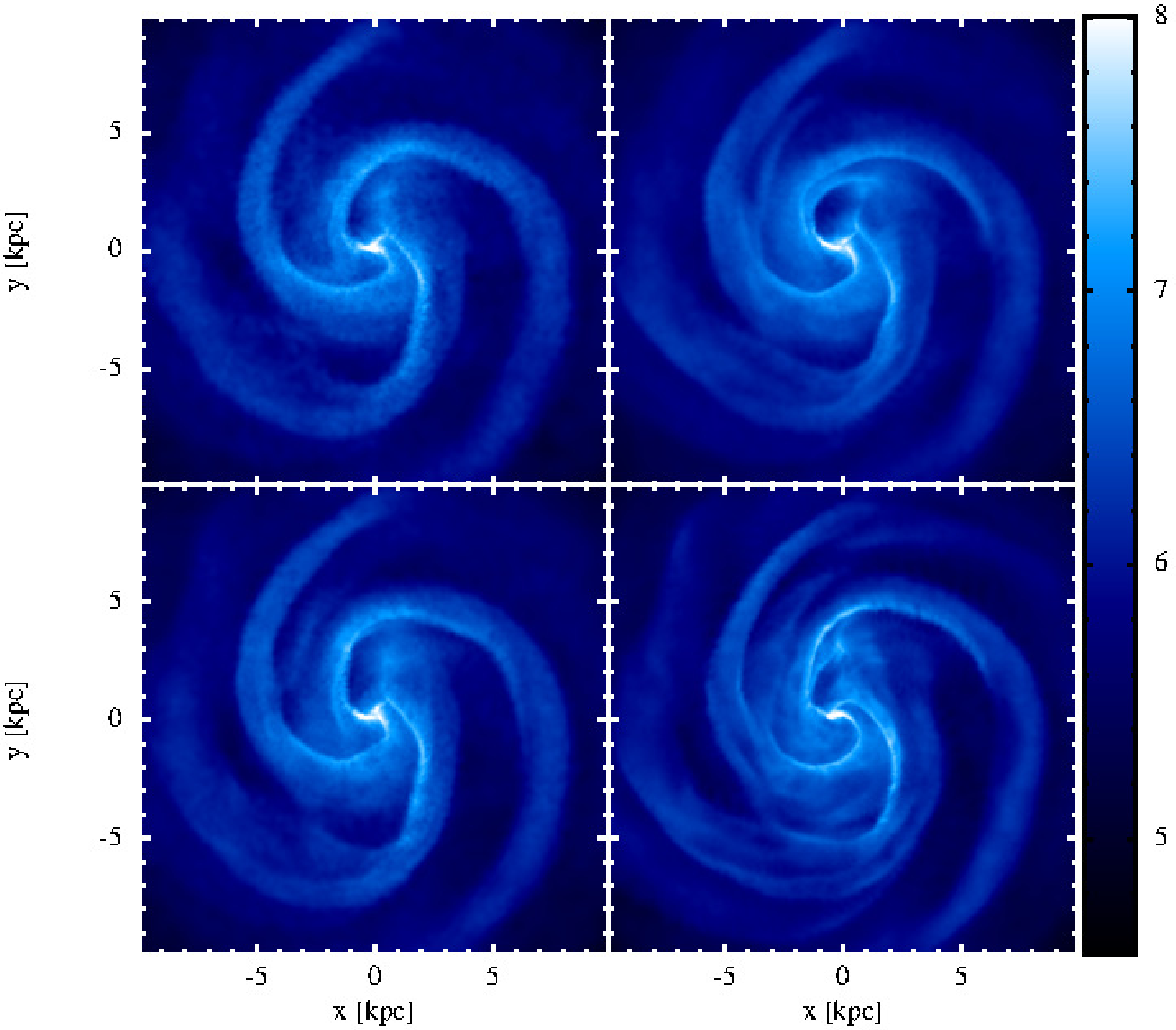

Figure 7. Face-on views of gas surface density for low resolution simulations with different values of α at t=1 Gyr: α=0.4 in the left upper panel, α=1.6 in the right upper panel, α=0.8 in the left lower panel and low resolution simulation using GIZMO in the right lower panel. The logarithmic density scale is in units of $M_\odot$ kpc^{-2}.

structures. In this section we focus on two times, just before (0.8 Gyr) and right at the beginning (1.1 Gyr) of the major gas inflow event. The properties of the stellar and gaseous distribution at the two times are highlighted in the upper and lower panels of figure 10.

The left panels represent the surface density contrast for the stars within the inner 3 kpc, defined as:

$$\delta_{\Sigma_\star}(R,\phi) = \frac{\Sigma_\star(R,\phi) - \langle\Sigma_\star(R,\phi)\rangle_\phi}{\langle\Sigma_\star(R,\phi)\rangle_\phi}, \tag{9}$$

where R and ϕ are the radial and azimuthal coordinates on the disc equatorial plane, $\Sigma_\star(R,\phi)$ is the stellar surface density and $\langle\Sigma_\star(R,\phi)\rangle_\phi$ is the average stellar surface density evaluated in annuli. The central panels show the gas density contrast

$$\delta_{\rho,\mathrm{gas}}(R,\phi) = \frac{\rho_{\mathrm{gas}}(R,\phi) - \langle\rho_{\mathrm{gas}}(R,\phi)\rangle_\phi}{\langle\rho_{\mathrm{gas}}(R,\phi)\rangle_\phi}, \tag{10}$$

evaluated on the disc mid-plane. The right panels show the intensity of the radial motions in km s^{-1}.

Before the major inflow event (upper left panel) a single three arm spiral structure is visible down to scales of about 300 pc. The gas is affected by the stellar non-axisymmetric structure and develops shocks at the edge of the stellar spirals, as observable in density contrast map (upper central panel). Clear shock fronts develop in the gas distribution, the gas dynamics is perturbed and radial motions are triggered (upper right panel).

A different picture is present at the triggering of the strong gas inflow episode (lower panels). At $t = 1.1$ Gyr the inner part (within $\approx$ 1 kpc) of the stellar three arm spiral structure decouples from the outer spiral structure, still evident at large scales ($\sim$ 3 kpc), as clearly visible in the lower left panel. Such decoupled structure is clearly observable in the gas density (lower central panel) map. The interplay between the outer and inner spiral structure increases the

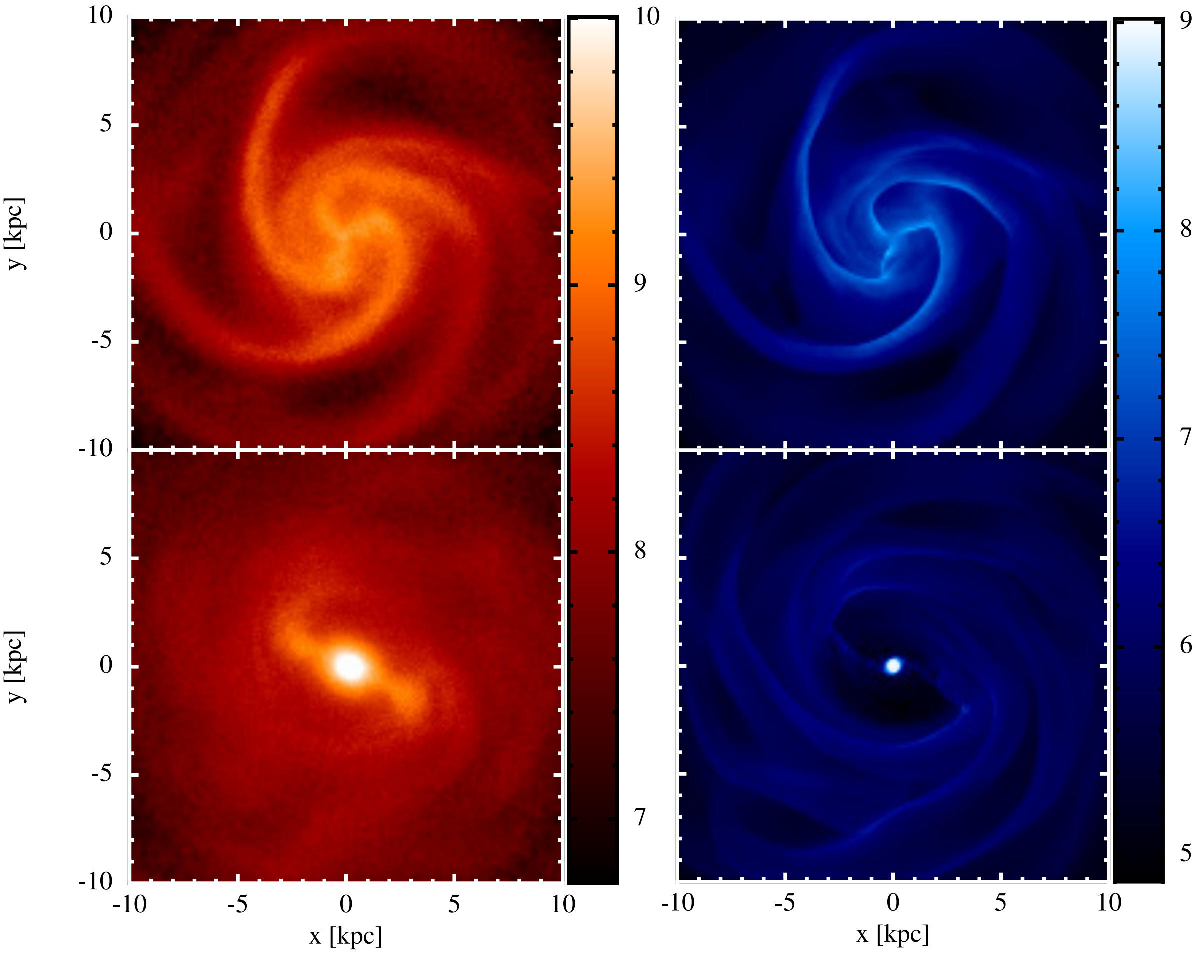

Figure 8. Right panels: logarithmic face-on views of the stellar surface density (in unit of M$_\odot$ kpc^{-2}) for the high resolution simulation (HR) at $t = 1$ Gyr (upper panel) and $t = 2.5$ Gyr (lower panel). Left panels: same as right panels for the gas surface density.

radial velocity of the gas in the central regions as well as the region participating to the radial inflow (lower right panel). The effect of the inner spiral decoupling onto the gas is reminiscent of the bars-within-bars scenario, originally proposed by Shlosman, Frank & Begelman (1989), in its *"stuff-within-stuff"* version (Hopkins & Quataert 2010), where the gravitational torques acting onto the gas are caused by non-axisymmetric structures not necessarily bar-like.

By the time a clear bar forms, all the gas affected by the nuclear spirals formed the central nuclear knot.

As a final comment, we stress that the nuclear regions of our galaxy do not show any evidence of gaseous clumps, whose migration could, in principle, cause the major gas inflow event (e.g. Bournaud, Elmegreen & Elmegreen 2007; Elmegreen, Bournaud & Elmegreen 2008). As a matter of fact, the further inwards one goes, the less evidence one has for clump formation. Such trend is not unexpected in systems with small gas-to-stellar mass fraction as the one we study here, that remains locally stable throughout the whole duration of the runs.

3.3.3 Nuclear disc

We devote the last part of our analysis to the structure of the gas nuclear structure forming during the major inflow event. In particular we will focus on the gas properties well after the nuclear structure formed and reached a stable configuration.

Figure 11 shows the density contrast of gas in the inner 3 kpc (left panel) and in the inner 400 pc (middle panel), and the radial velocity map (right panel) at about 2.5 Gyr. The orientation of the bar is traced by the inflowing streams of gas that connect the outer regions of the galaxy with the inner gaseous structure. The gas in the inner few hundreds of pc forms a rotating disc. Within the disc nuclear spirals are observable down to few tens of pc, traced by local gaseous overdensities (lighter regions in the central panel) corresponding to inflowing gas (blue and green regions in the right panel). A careful analysis of the stellar distribution does not show any central structure (neither in the form of spirals nor of bar). We therefore interpret the central two armed spirals as the effect that the outer bar has onto the gas within its ILR, as discussed analytically in Maciejewski (2004a) and observed in numerical simulations of the response of gas to a bar-like analytical potential (Maciejewski 2004b).

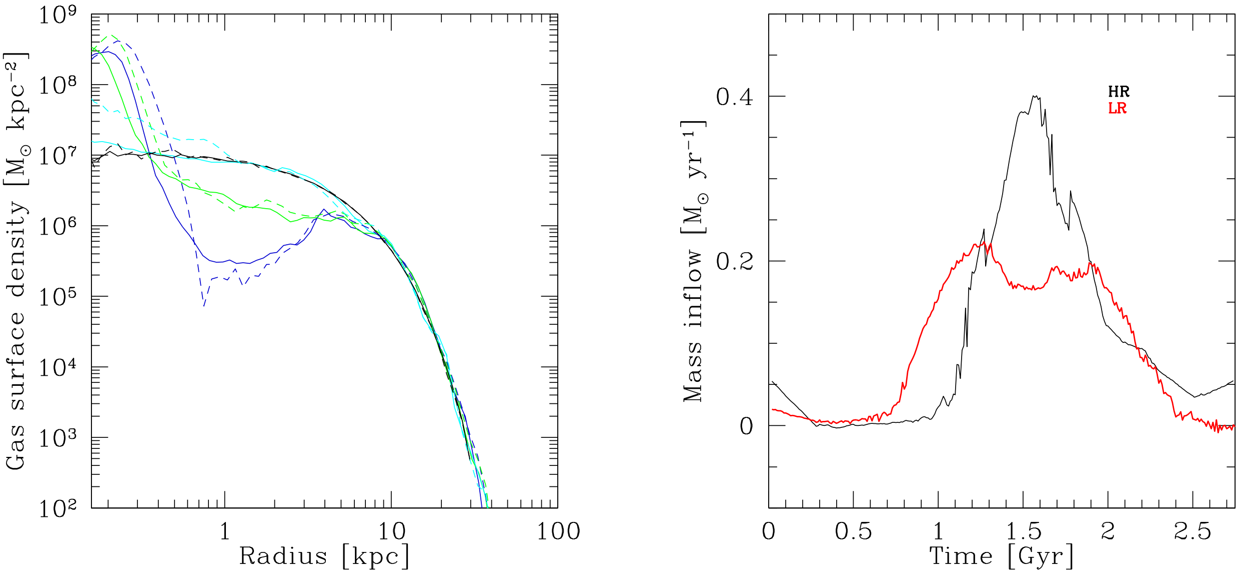

Figure 9. Left panel: surface density profile of the gas. Solid lines refer to the highest resolution run HR. The results of the low resolution run LR are reported with dashed lines for comparison. Black, cyan, green and blue curves refer to t =0,1,2, and 3 Gyr, respectively. Right panel: gaseous mass inflow as a function of time. Black and red lines represent the inflow rate computed at 0.3 kpc from the center in run HR and LR respectively.

As a note of caution we stress that the accretion rate at late times and, more in general, the long-term evolution of the central gaseous concentration would be significantly different if the gas would be allowed to form stars, possibly forming a pseudo-bulge like structure. We plan to study the possible effect of star formation on the nuclear scale gas dynamics in a future investigation.

4 DISCUSSION AND CONCLUSIONS

In this paper we studied the gas response to the formation and evolution of a stellar bar in an isolated disc galaxy. The galaxy is initially unstable to the formation of non-axisymmetric structures and develops multiple spiral arms in the first Gyr, that evolve in a central stellar bar at $t \sim 2$ Gyr. The forming bar slows down with time, and buckles in its central $\gtrsim$ kpc region.

During the first spiral arm dominated phase the gas, forced by the stellar potential, forms clearly defined spiral arms. During this phase a major episode of gas inflow takes place, larger by a factor of $\gtrsim 3$ than any other inflow event after the bar formation. The analysis of the higher resolution simulation shows that the trigger of the major inflow is the decoupling of the nuclear regions of the three armed spiral from the outer counterpart. At later times, when the stellar bar is already established, a low gas density annulus (here defined as the dead zone) between the bar corotational and the inner Lindblad resonances $R_{\rm ILR} \lesssim R \lesssim R_{\rm C}$ is clearly observable in the simulations. We notice that such a gas depleted region is often observed in local samples of barred spiral galaxies, as extensively discussed in Gavazzi et al. (2015).

We checked our results against the numerical viscosity used, and we demonstrated that the gas dynamics is little affected by the exact value of the viscosity parameter in the SPH runs, and by the exact hydrodynamical treatment of the gas. We also studied the dependence of our results on the numerical resolution. We found that, although the qualitative evolution of the gas is resolution independent, the exact time at which the non axisymmetric structures develop and the actual maximum inflow rate at small (but completely resolved) scales do depend on the resolution achieved. As discussed above, the difference in the timescales for the inflow and for the bar formation are probably due to a lower shot noise in the highest resolution initial conditions. The difference in the magnitude of the maximum inflow rate, instead, is due to the fact that the bar itself as well as all the non-axisymmetric structure are better resolved in the highest resolution run, resulting in a more effective torquing of the gas.

Independently of the exact numerical implementation, we find that the flux of gas reaching the most central regions of the galaxy peaks during the bar formation phase, and not when the bar is fully established. The explanation of such result is twofold: (1) since bars are quite efficient in driving the gas within their corotational radius toward the centre, after few bar orbits the central region of the galaxy is mostly gas free (as already noted by Berentzen et al. 1998), and there is no remaining gas to be torqued by the bar; (2) in our simulations the forming bar slows down as the galaxy evolves, increasing its ILR and corotational radii (in agreement with, e.g. Sellwood 1981; Combes & Sanders 1981; Halle et al. 2015). As a consequence, the gas that is perturbed by the early fast-precessing bar reaches regions significantly more nuclear than gas perturbed at later times.

The low efficiency of large, long-lived and easy to spot bars in fueling the very central regions of galaxies can explain why many observational studies do not find significant links between bars and AGN activity. The high efficiency of

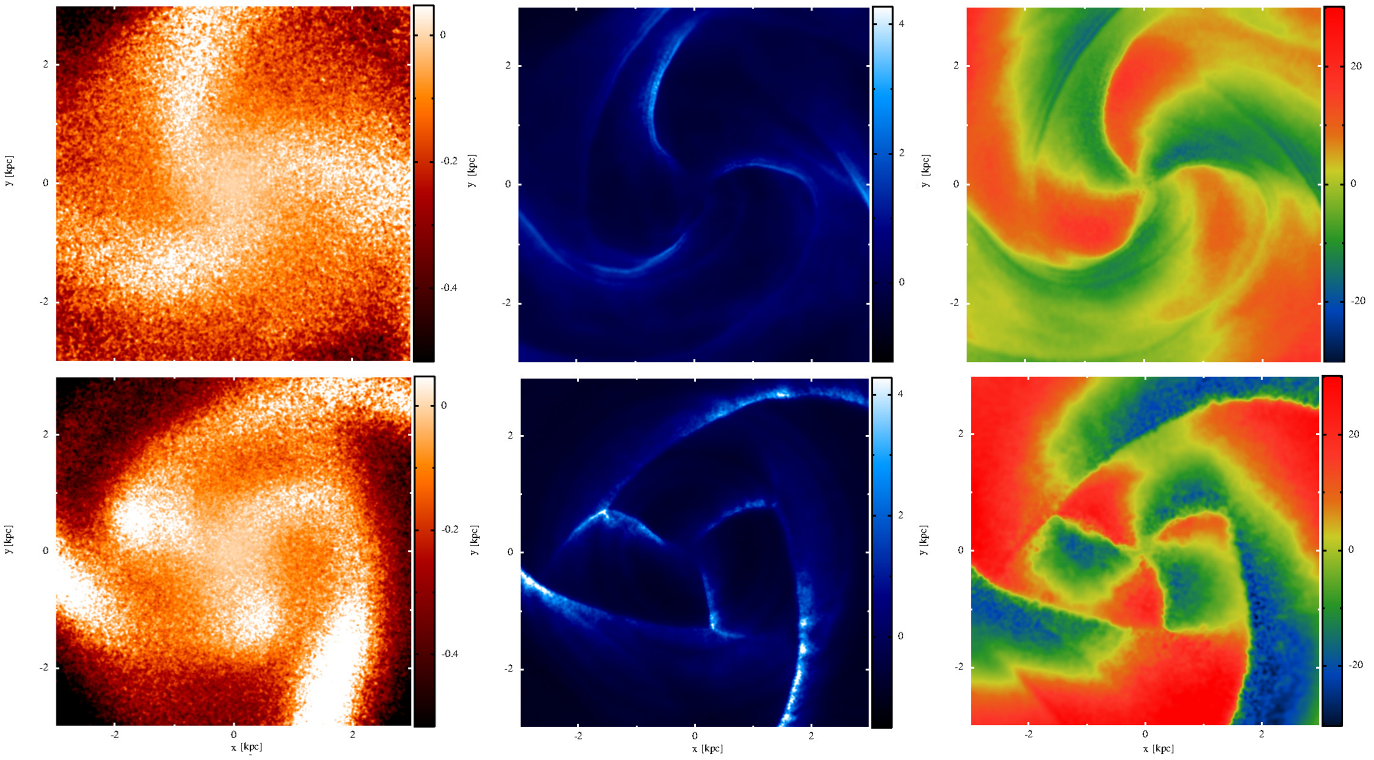

Figure 10. Upper panels: stellar surface density contrast (left), gas density contrast (middle) and radial velocity map (right, in unit of km s^{-1}) for gas in the inner 3 kpc at t =0.8 Gyr. Lower panels: same as the upper panels at t = 1.1 Gyr. See text for details.

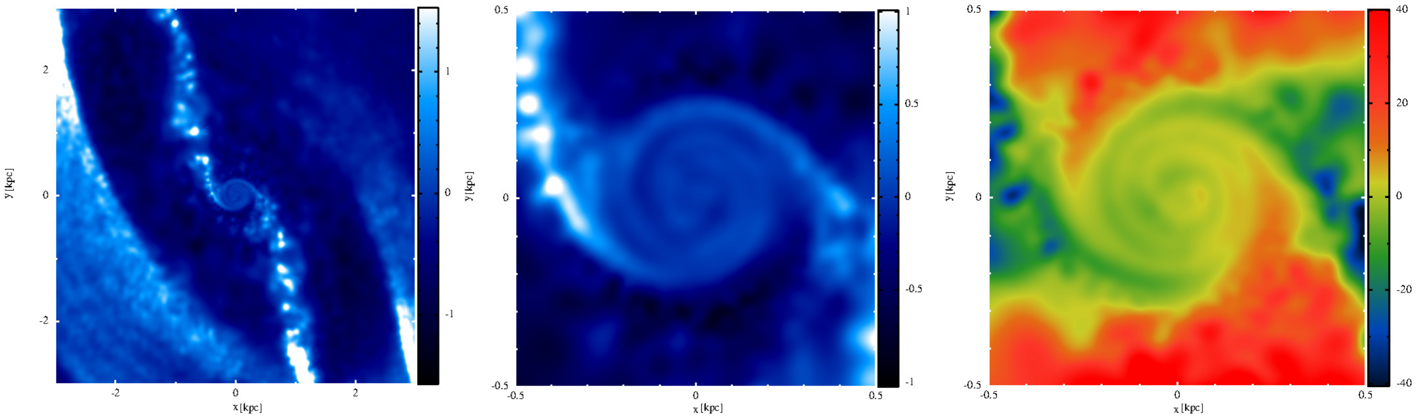

Figure 11. Gas density contrast in the inner 3 kpc (left panel) and 500 pc (central panel). The right panel shows the radial velocity for gas (in unit of km s^{-1}) in the inner 500 pc at 2.5 Gyrs. See text for details.

the *bar formation* process in driving strong inflows toward the very central region of galaxies hints, on the other hand, at a possibly underestimated importance of bar driven AGN activity in disc galaxies.

Finally, the analysis of the long lived nuclear gaseous disc shows that the outer, large scale, bar keeps on exerting a dynamical effect onto the gas within few hundreds of pc. A two armed spiral can be observed both in the gas density distribution and in the gas dynamics, as already discussed by Maciejewski (2004a) and Maciejewski (2004b).

As a final note of caution, we highlight the main shortcoming of the simulation suite discussed here:

(i) Our simulations lack physically motivated prescriptions for gas radiative cooling, star-formation and any star-formation-related feedback, as well as accretion onto a possibly present massive black hole and the related AGN feedback.We stressed that the dynamics of the gas in the region studied here (down to few hundred pc from the galaxy centre) is not strongly affected by the lack of additional physics, as the gas inflow happens on few orbital timescales. As a matter of fact, we regard the lack of additional physics as a plus in our runs, as it allows to clearly highlight the dynamical processes ongoing in the simulation in a controlled system. On the other hand, the lack of star formation and related feedbacks does not allow us to draw firm conclusions about the long term evolution of the gas at small scales, as a significant fraction of the gas could turn into stars on the Gyr timescales of the simulations. A follow up set of runs including additional physics is currently in preparation;

(ii) All the simulations discussed here share the same idealized initial conditions. We regard this as the main drawback of our study. Because of this we cannot use our runs to make any general prediction about barred galaxies in general. We stress, however, that our simple runs do highlight the possible relevance of early gas inflow during the bar formation phase. We plan to check our results with fully evolving isolated galaxy simulations starting from different initial conditions and, with a considerable increase of the computational cost, with cosmological simulations.

ACKNOWLEDGMENTS

We thank the anonymous Referee for her/his suggestions that significantly improved the quality of the paper. We acknoledge Alessandro Lupi for the help in the technical aspects of the runs, and for his comments on the paper. We further thank Silvia Bonoli, Pedro R. Capelo, Guido Consolandi, Jorge Cuadra, Roberto Decarli, and Giuseppe Gavazzi for their comments and insights.

References

Alonso M. S., Coldwell G., Lambas D. G., 2013, A&A, 549, A141
Athanassoula E., 1992, MNRAS, 259, 345
Balsara D. S., 1995, Journal of Computational Physics, 121, 357
Barnes J., Hut P., 1986, Nature, 324, 446
Berentzen I., Heller C. H., Shlosman I., Fricke K. J., 1998, MNRAS, 300, 49
Berentzen I., Shlosman I., Martinez-Valpuesta I., Heller C. H., 2007, ApJ, 666, 189
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press
Bournaud F., Elmegreen B. G., Elmegreen D. M., 2007, ApJ, 670, 237
Cheung E. et al., 2015, MNRAS, 447, 506
Cisternas M. et al., 2013, ApJ, 776, 50
Cole D. R., Debattista V. P., Erwin P., Earp S. W. F., Roškar R., 2014, MNRAS, 445, 3352
Combes F., Sanders R. H., 1981, A&A, 96, 164
Elmegreen B. G., Bournaud F., Elmegreen D. M., 2008, ApJ, 688, 67
Fiacconi D., Feldmann R., Mayer L., 2015, MNRAS, 446, 1957
Gavazzi G. et al., 2015, arXiv:1505.07836
Goz D., Monaco P., Murante G., Curir A., 2014, arXiv:1412.2883
Halle A., Di Matteo P., Haywood M., Combes F., 2015, arXiv:1501.00664
Hernquist L., 1990, ApJ, 356, 359
Hernquist L., 1993, ApJS, 86, 389
Hinshaw G. et al., 2013, ApJS, 208, 19
Ho L. C., Filippenko A. V., Sargent W. L. W., 1997, ApJ, 487, 591
Hopkins P. F., 2014, arXiv:1409.7395
Hopkins P. F., Quataert E., 2010, MNRAS, 407, 1529
Hunt L. K., Malkan M. A., 1999, ApJ, 516, 660
Jogee S., Scoville N., Kenney J. D. P., 2005, ApJ, 630, 837
Kim W.-T., Seo W.-Y., Kim Y., 2012, ApJ, 758, 14
Knapen J. H., Shlosman I., Peletier R. F., 2000, ApJ, 529, 93
Komatsu E. et al., 2011, ApJS, 192, 18
Kormendy J., 2013, Secular Evolution in Disk Galaxies, Falcón-Barroso J., Knapen J. H., eds., p. 1
Kraljic K., Bournaud F., Martig M., 2012, ApJ, 757, 60
Laine S., Shlosman I., Knapen J. H., Peletier R. F., 2002, ApJ, 567, 97
Laurikainen E., Salo H., Buta R., 2004, ApJ, 607, 103
Lee G.-H., Park C., Lee M. G., Choi Y.-Y., 2012a, ApJ, 745, 125
Lee G.-H., Woo J.-H., Lee M. G., Hwang H. S., Lee J. C., Sohn J., Lee J. H., 2012b, ApJ, 750, 141
Lodato G., Price D. J., 2010, MNRAS, 405, 1212
Maciejewski W., 2004a, MNRAS, 354, 883
Maciejewski W., 2004b, MNRAS, 354, 892
Martinet L., Friedli D., 1997, A&A, 323, 363
Mayer L., Wadsley J., 2004, MNRAS, 347, 277
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
Monaghan J. J., 1992, ARA&A, 30, 543
Monaghan J. J., 1997, Journal of Computational Physics, 136, 298
Mulchaey J. S., Regan M. W., 1997, ApJ, 482, L135
Murray J. R., 1996, MNRAS, 279, 402
Nair P. B., Abraham R. G., 2010, ApJ, 714, L260
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Oh S., Oh K., Yi S. K., 2012, ApJS, 198, 4
Regan M. W., Sheth K., Vogel S. N., 1999, ApJ, 526, 97
Regan M. W., Teuben P. J., 2004, ApJ, 600, 595
Roberts, Jr. W. W., Huntley J. M., van Albada G. D., 1979, ApJ, 233, 67
Romano-Díaz E., Shlosman I., Heller C., Hoffman Y., 2008, ApJ, 687, L13
Sanders R. H., Huntley J. M., 1976, ApJ, 209, 53
Scannapieco C., Athanassoula E., 2012, MNRAS, 425, L10
Sellwood J. A., 1981, A&A, 99, 362
Sellwood J. A., 2014, Reviews of Modern Physics, 86, 1
Shlosman I., Frank J., Begelman M. C., 1989, Nature, 338, 45
Springel V., 2005, MNRAS, 364, 1105
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Springel V., Hernquist L., 2002, MNRAS, 333, 649
Villa-Vargas J., Shlosman I., Heller C., 2010, ApJ, 719, 1470
Wada K., 2001, ApJ, 559, L41
Wada K., Norman C. A., 2001, ApJ, 547, 172

Mon. Not. R. Astron. Soc. **000**, 1–12 (2014) Printed 25 May 2016 (MN LaTeX style file v2.2)

Growing massive black holes through super-critical accretion of stellar-mass seeds

A. Lupi[1,3], F. Haardt[1,3], M. Dotti[2,3], D.Fiacconi[4], L. Mayer[4] & P. Madau[4,5]
[1] *DiSAT, Università degli Studi dell'Insubria, Via Valleggio 11, I-22100 Como, Italy*
[2] *Dipartimento di Fisica, Università degli Studi di Milano Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[3] *INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[4] *Center for Theoretical Astrophysics and Cosmology, Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*
[5] *Department of Astronomy & Astrophysics, University of California, 1156 High Street, Santa Cruz, CA 95064, USA*

Draft 25 May 2016

ABSTRACT

The rapid assembly of the massive black holes that power the luminous quasars observed at $z \sim 6-7$ remains a puzzle. Various direct collapse models have been proposed to head-start black hole growth from initial seeds with masses $\sim 10^5 \mathrm{M}_\odot$, which can then reach a billion solar mass while accreting at the Eddington limit. Here we propose an alternative scenario based on radiatively inefficient super-critical accretion of stellar-mass holes embedded in the gaseous circum–nuclear discs (CNDs) expected to exist in the cores of high redshift galaxies. Our sub-pc resolution hydrodynamical simulations show that stellar-mass holes orbiting within the central 100 pc of the CND bind to very high density gas clumps that arise from the fragmentation of the surrounding gas. Owing to the large reservoir of dense cold gas available, a stellar-mass black hole allowed to grow at super-Eddington rates according to the "slim disc" solution can increase its mass by 3 orders of magnitudes within a few million years. These findings are supported by simulations run with two different hydro codes, RAMSES based on the Adaptive Mesh Refinement technique and GIZMO based on a new Lagrangian Godunov-type method, and with similar, but not identical, sub-grid recipes for star formation, supernova feedback, black hole accretion and feedback. The low radiative efficiency of super-critical accretion flows are instrumental to the rapid mass growth of our black holes, as they imply modest radiative heating of the surrounding nuclear environment.

Key words: black hole - galaxy formation - galaxy evolution.

1 INTRODUCTION

Observations of luminous quasars at very high redshift pose crucial questions on the formation of massive black holes (MBHs) along the history of the Universe. The most distant quasar to date, ULAS J1120+0641, lies at redshift $z = 7.084$, and it is believed to be powered by a $\sim 2 \times 10^9\,\mathrm{M}_\odot$ MBH (Mortlock et al. 2011) that was therefore in place (and shining) 0.76 Gyr after the Big Bang. Together with the handful of bright quasars observed by the *Sloan Digital Sky Survey* at $z \gtrsim 6$ (Fan et al. 2006), ULAS J1120+0641 sets tight constraints on any model for the formation and growth of MBHs at early epochs.

It has been long thought that the first MBH seeds were *light*, specifically $\sim 100\,\mathrm{M}_\odot$ remnants of the first generation of stars, plausibly formed at $z \gtrsim 20$ (e.g., Madau & Rees 2001; Haiman & Loeb 2001; Heger et al. 2003; Volonteri, Haardt & Madau 2003; Madau et al. 2004). If gas accretion is Eddington-limited, light seeds could grow to the supermassive variety by $z \sim 7$ only if (i) gas accretion continued unimpeded at the Eddington rate for $\gtrsim 0.6$ Gyr, and if (ii) the mass-to-light conversion efficiency of the accretion process was not high, $\epsilon \lesssim 0.1$ (Tanaka & Haiman 2009). The first condition seems hard to satisfy in the shallow potential of low–mass dark matter haloes, as radiative feedback from the progenitor and from BH accretion itself dramatically affects the gas inflow and its supply to the hole, resulting in sub–Eddington rates, therefore negligible mass growth (e.g., Wise, Turk & Abel 2008; Milosavljević et al. 2009; Alvarez, Wise & Abel 2009). The second condition is problematic too, as it requires a radiative efficiency well below that proper of accretion onto rapidly rotating black holes. Indeed, there are mounting evidences that the most massive holes at high redshifts power radio–loud AGNs (see, e.g., Ghisellini et al. 2014). These are thought to be associated with Kerr holes – though observational evidences of the widely accepted jet–spin connection are, at best, scarce, even in the well studied Galactic stellar black hole candidates (see, e.g., Russell, Gallo & Fender 2013).

Over the last decade, a number of alternatives to the "light seed scenario" have been proposed. *Heavy* seeds, with masses as large as $\gtrsim 10^4\,\mathrm{M}_\odot$, may form through the direct collapse of low angular momentum gas at high redshifts (see, e.g., Loeb & Rasio 1994; Bromm & Loeb 2003; Koushiappas, Bullock & Dekel 2004; Spaans & Silk 2006; Mayer et al. 2010, 2014), likely via the intermediate stages of a supermassive star and a "quasistar" (Begelman, Volonteri & Rees 2006; Begelman, Rossi & Armitage 2008; Begelman 2010; Dotan, Rossi & Shaviv 2011). For such mechanism to work one needs to, at the same time (i) avoid fragmentation, (ii) effectively dissipate angular momentum, and (iii) drive gas towards the centre of the protogalaxies at a rate $\gtrsim 1\,\mathrm{M}_\odot$ yr^{-1} (e.g, Ferrara, Haardt & Salvaterra 2013; Latif et al. 2015). There is still no consensus, to date, about where and when such conditions are actually fulfilled.

In a previous paper (Madau, Haardt & Dotti 2014, hereinafter MHD14), we discussed super–critical (i.e., super–Eddington) accretion onto stellar mass seeds as a possible mechanism for bypassing the above difficulties. We used the radiatively–inefficient "slim–disc" solution (Abramowicz et al. 1988) – advective, optically thick flows that generalise the standard Shakura & Sunyaev solution (Shakura & Sunyaev 1973) – and showed how mildly super–Eddington accretion significantly eases the problem of assembling MBHs in less than a billion year. Because of the (accretion–rate dependent) low radiative efficiencies of slim discs around non–rotating *as well as* rapidly rotating holes, the accretion time–scale in this regime is almost independent of the spin parameter. It is this unique feature of slim discs that makes such models so appealing.

In MHD14 (see also Volonteri, Silk & Dubus 2015) we briefly discussed how conditions for super–critical accretion are physically plausible in the dense environment of high redshift massive protogalaxies. Here, we elaborate upon this concept by means of high resolution simulations of a cluster of stellar mass black holes orbiting the central ~ 200 pc of a gas–rich galaxy. We show how the interplay between gas dynamics and the black holes can easily lead to the formation of a MBH in the centre of the system within few million years. Though our simulations are highly idealised and should be thought as a proof-of-concept of the scenario we are proposing, they highlight the basic point, i.e. that super–Eddington accretion in well–formed, evolved galaxies is an attractive route to the formation of massive black holes. In fact, a population of stellar mass black holes is expected to reside in the inner ~ 200 pc, the circum–nuclear disc can provide enough gas to be accreted, and negative feedback is negligible in the high–density clumps developed in the disc.

This is the first of a series of papers devoted at the study of the effect of a radiatively inefficient BH feedback on the early growth of stellar mass BHs embedded in a circum–nuclear gas disc. In a forthcoming paper (Fiacconi et al., in preparation) we will discuss the nature and properties of the circum–nuclear gas by means of a fully cosmological, high–resolution simulation of the formation of a gas–rich massive disc galaxy at $z > 6$.

2 SIMULATIONS

We perform a suite of 6 simulations, using the adaptive mesh refinement (AMR) code RAMSES (Teyssier 2002) ("_R" runs) and the new mesh–free code GIZMO (Hopkins 2014) ("_G" runs), which implements a new Lagrangian method to solve the hydrodynamic equations (similar to Godunov–type schemes). The two codes are equipped with similar, but not identical sub-grid recipes for star formation, supernova feedback, BH accretion and BH feedback. While RAMSES is a well-tested AMR code which has been already successfully employed to study circum–nuclear gas discs in merger remnants (Chapon, Mayer & Teyssier 2013), GIZMO employs a novel numerical technique which is in principle well suited to study the flow in a highly dynamical situation while conserving angular momentum and limiting numerical diffusion, with advection errors that are smaller than grid-based cartesian AMR codes such as RAMSES (Hopkins 2014). GIZMO is thus supposed to combine the strengths of grid-based and SPH codes while retaining the tree-based gravity solver inherited from GADGET3, which guarantees high accuracy as well as fast computation for self-gravitating discs like those under study here. The use of two different powerful numerical techniques is aimed at checking the robustness and reproducibility of our results.

2.1 Initial conditions

In our simulations we consider the nuclear region of a high–redshift massive spiral galaxy, where we assume a gaseous circum–nuclear disc (CND hereafter) is hosted. We model the CND following an exponential surface density profile (see Fig. 1, top panel), with total mass $10^8\,\mathrm{M}_\odot$ and scale radius 50 pc. The disc is embedded in a stellar spherical background following an Hernquist profile, with scale radius 100 pc and total mass $2\times10^8\,\mathrm{M}_\odot$. Note that the adopted gas mass from the CND is at the lower end of that of CNDs observed in low z merger remnants (e.g., Medling et al. 2014), and should thus be regarded as a conservative choice for the purpose of our paper. At high redshift galaxies are indeed expected to be more gas rich, which is supported by both theoretical arguments and observational evidence.

The disc, modelled as an ideal gas with polytropic index $\gamma = 5/3$, is set in hydrostatic equilibrium in the global potential with an initial temperature $T = 10^4$ K. The initial conditions are produced using the public code GD_BASIC, described in Lupi, Haardt & Dotti (2015). We show in Fig. 1, bottom panel, the velocity profile of the gas component in the disc. The CND is Toomre unstable since the beginning of the run in the region between 10 pc and 70 pc, as reported in Figure 2 (blue solid line). In order to check if such low temperature could result in overestimated clumpiness we perform two runs of the disc in isolation, one starting with the above mentioned temperature and one with $T = 3\times10^4$ K (the equilibrium temperature in the case of a disc heated by UV background radiation at $z\sim 8$). We do not include UV background radiation in the simulation, assuming the disc will become self–shielded. Despite the initially higher

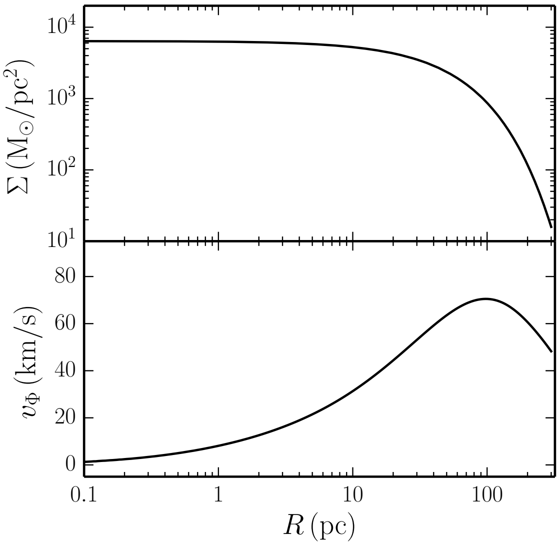

Figure 1. Initial surface density (top panel) and circular velocity (bottom panel) profiles of the circum–nuclear disc.

temperature, we find that after only few 10^4 years the disc temperature drops to 7×10^3 K in both runs, converging to an identical profile for the Toomre parameter Q (Figure 2, dashed lines).

We then assume that previous star-formation episodes left a population of stellar mass black holes in the galaxy nucleus. The mass of such "black hole seeds" $M_{\rm BH}$ is alternatively set 20 or $100\mathrm{M_\odot}$. We initially distribute the BHs uniformly within the inner 150 pc of the CND. The BHs lay in the disc plane and have an initial velocity equal to the local circular velocity. We add a randomly oriented velocity component sampled from a normal distribution with standard deviation $\sigma \sim 20\%$ of the maximum circular velocity.

2.2 RAMSES Eulerian simulations

We perform two simulations with RAMSES at two different spatial resolutions, namely 0.4 pc ("low" runs) and 0.1 pc ("med" runs). The mass resolution is $10^3\,\mathrm{M_\odot}$ at the quasi–Lagrangian threshold for refinement. We include the radiative cooling of the gas adopting the standard prescriptions employed in the code (see Teyssier et al. 2013, for details). In order to prevent spurious fragmentation at the highest refinement level we add a polytropic pressure term to the gas component (described as a polytrope with $\gamma = 5/3$ and temperature 10^3 K at $2 \times 10^5\,\mathrm{cm^{-3}}$), ensuring to resolve the Jeans length with at least 4 cells at the highest refinement level.

We set a star formation density and temperature threshold of $\rho_{\rm thr} = 2 \times 10^5\ \mathrm{cm^{-3}}$ and $T_{\rm thr} = 2 \times 10^4$ K, and a typical star formation time–scale of 1 Myr. We also assume a time delay between star formation and the corresponding SNa explosion event of 1 Myr, with a SNa yield of 0.15 (corresponding to stars with masses above $8\,\mathrm{M_\odot}$ for a Salpeter IMF). In order to model non thermal processes associated with SNa events, we include the blast–wave like feedback described in Teyssier et al. (2013). In this feedback recipe the SNa energy budget is decoupled from the thermal energy of the gas, preventing, for a typical timescale $\simeq 20$ Myr, the gas to radiatively cool. In our runs we assume a primordial gas composition and we include subsequent metal pollution due to SNae.

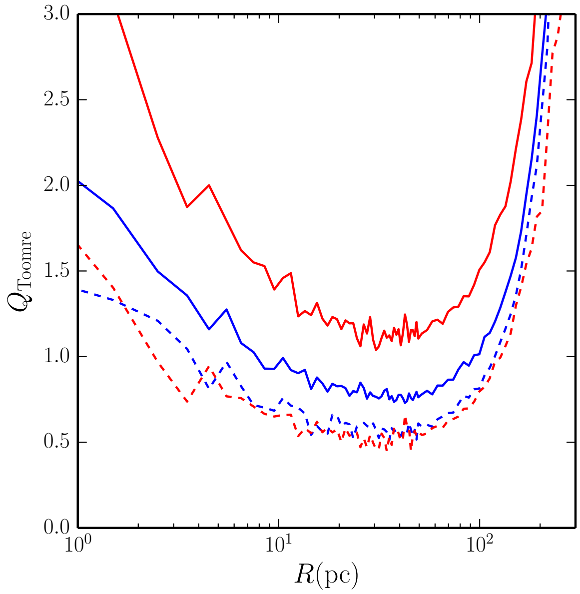

Figure 2. Toomre parameter for the isolated CNDs at $t = 0$ Myr (solid lines) and $t = 0.05$ Myr (dashed lines). The blue line corresponds to an initial temperature $T = 10^4$ K and the red one to an initial temperature $T = 3 \times 10^4$ K.

2.3 GIZMO Lagrangian simulations

Since the public version of GIZMO only includes basic hydrodynamics and gravity, we implement in the code the additional recipes necessary to model gas cooling, star formation, type II SNa feedback and gas accretion onto BHs, in a fashion similar to RAMSES implementations.

We include radiative cooling by means of the GRACKLE[1] chemistry and cooling library, which provides both equilibrium and non–equilibrium chemistry (The Enzo Collaboration et al. 2014; Kim et al. 2014). In our runs we employ the equilibrium cooling curve for primordial species (atomic H and He), and tabulated metal cooling and heating from Cloudy (Ferland et al. 2013).

Gas particles are eligible for star formation when they match the same criteria for density and temperature adopted in our RAMSES runs and belong to a converging flow (i.e., $\nabla \cdot \mathbf{v} < 0$). Resulting star particles are generated locally according to the Schmidt law (Kennicutt 1998), and using a stochastic prescription as described in Stinson et al. (2006). We model SNa feedback assuming that after a typical timescale of 1 Myr stars above $8\,\mathrm{M_\odot}$ explode as type II SNae, releasing $10^{50}\,\mathrm{erg/\,M_\odot}$ in the form of purely thermal energy.

---

[1] `http://grackle.readthedocs.org`

The non thermal processes that energise the SNa blast–wave have a typical dissipation timescale much longer than that of the thermal energy. Therefore, we implement a “delayed cooling” prescription as follows: we temporarily inhibit radiative cooling for gas particles within the SNa maximum extension radius $R_{\rm E}$ defined as in Chevalier (1974). The SNa energy is then distributed among the gas particles lying within $R_{\rm E}$, according to their distance from the SNa, weighted through the kernel function used in the code. This is different from the approach taken in the RAMSES runs, in which the energy is wholly released within the cell hosting the progenitor. In GIZMO runs we limit the cooling delay time to 5 Myr only (i.e., 4 times smaller than what assumed in our RAMSES runs). We checked that this set up provides consistent results between RAMSES and GIZMO feedback implementations.

We perform three simulations allowing for two different gravitational resolutions, i.e., 0.16 pc (“low” runs) and 0.02 pc (“high” run). We set the same gravitational resolution for gas particles and BH particles. We use 10^5 particles for the “low” runs and 10^7 for the “high” run, corresponding to a mass resolution of $10^3\,{\rm M}_\odot$ and $10\,{\rm M}_\odot$, respectively. In our runs we use the finite mass mode available in the code, in which mass transfer between particles is forbidden, so that our simulations are purely Lagrangian.

2.4 BH accretion

In RAMSES runs we evaluate the accretion onto the stellar mass BHs using the so-called “flux accretion” prescription. In such a scheme the accretion rate is the mass flux rate within the BH accretion zone, which consists of a sphere of radius equal to 4 cells, i.e., $\dot{M}_{\rm acc} = \int -\nabla\cdot[\rho\Delta\mathbf{v}]d^3\mathbf{x}$, where the integral is over the volume of the accretion zone and $\Delta\mathbf{v}$ is the gas–BH relative velocity (see Bleuler & Teyssier 2014, for a detailed description of the implementation).

In order to get a more accurate BH dynamics and to best resolve the accretion rate, we force the region near to each BH to always be at the maximum refinement level, as described in Lupi, Haardt & Dotti (2015). Forcing the resolution close to the BHs at the highest possible level guarantees that nearby cells own a mass $\lesssim 5\times M_{\rm BH}$ during the whole BH accretion history. This allows us to set $20\,{\rm M}_\odot$ as the initial mass of the BHs in our RAMSES runs.

In GIZMO runs we allow gas accretion onto the stellar–mass BHs for particles lying within a distance from the hole than encompasses an effective number of gas particles $N = 32$. Such *kernel radius h* is therefore implicitly defined through the following relation:

$$\frac{4}{3}\pi h^3 \sum_j W(z_j) = 32. \tag{1}$$

The *volume partition kernel* $W(z_j)$ is a function of the normalised distance to the BH $z_j \equiv |\mathbf{x}_j - \mathbf{x}_{\rm BH}|/h$, where $\mathbf{x}_j$ and $\mathbf{x}_{\rm BH}$ are the position vectors of the j-th gas particle and of the accreting BH, respectively (for further details see Hopkins 2014). Since the kernel radius strongly depends upon the gas density, our scheme would typically overestimate the accretion rate when a BH moves in a low density medium. In order to overcome this problem we enable accretion onto a BH only when the kernel radius is $\leqslant 10$ times the softening length of the sink particle. The accretion rate is then computed as described for RAMSES runs.

Because of the large mass of gas particles the dynamics of BHs as light as 20 ${\rm M}_\odot$ cannot be properly solved for. Therefore, we start from a larger initial BH mass, i.e., $M_{\rm BH} = 100\ {\rm M}_\odot$. With such a choice, BHs in the “high” runs have resolved dynamics since the very beginning of the simulation. In the “low” case, the initial BH dynamics and growth is instead affected by the lack of mass resolution. However, as will be discussed in the next section, some BHs grow above 1000 ${\rm M}_\odot$ in a very short time, making dynamics quickly reliable.

With respect to the standard Bondi-Hoyle model, this recipe does not make any geometrical assumption for the gas flow, allowing for a more accurate estimation of the accretion rate, where the effect of angular momentum on the resolved scales is taken into account. However, despite the high resolution reached with the “high_G” run, we are unable to properly follow the gas from sub–parsec scales down to the accretion disc scale. This resolution limit could result in a overestimated and more efficient accretion. Such a convergence study is beyond the scope of the present work.

2.5 BH feedback

The public release of RAMSES already provides a prescription for AGN thermal feedback (Dubois et al. 2014), in which the radiation produced by accretion is stored until the total budget is large enough to heat up the surrounding gas to at least 10^7 K. This prescription has been initially proposed by Booth & Schaye (2009) to prevent the gas from immediately loose the small amount of additional thermal energy gained after each time–step, which would result in an ineffective feedback. The RAMSES prescription assumes a fixed accretion radiative efficiency $\epsilon = 0.1$, and a fixed fraction $= 15\%$ of the accretion energy to be released to gas. In our simulations we suitably modify this recipe to include the effects of accretion in the fashion of slim disc (Sądowski et al. 2014). To this aim, we estimate ϵ using the analytical fit to the numerical results by Sądowski et al. (2014) provided by Madau, Haardt & Dotti (2014):

$$\epsilon = \frac{r}{16}A(a)\left[\frac{0.985}{r+B(a)} + \frac{0.015}{r+C(a)}\right], \tag{2}$$

where $r = \dot{M}_{\rm E}/\dot{M}$. Here $\dot{M}_{\rm E} = 16L_{\rm E}/c^2$ where $L_{\rm E}$ is the Eddington luminosity. A, B, C are fitting functions scaling with the BH spin a as

$$A(a) = (0.9663 - 0.9292a)^{-0.5639}, \tag{3}$$

$$B(a) = (4.627 - 4.445a)^{-0.5524}, \tag{4}$$

$$C(a) = (827.3 - 718.1a)^{-0.7060}. \tag{5}$$

At each accretion event we compute the released energy allowed to feedback on nearby particles using this new value for ϵ instead of the fixed value 0.1, while the spin is always fixed at $a = 0.99$ for all BHs. We implement in GIZMO the same prescription for BH feedback. In all simulations we do not include other possible forms of BH feedback, e.g., momentum-driven.

Finally, in order to check whether super–critical accretion is instrumental in leading to very large $M_{\rm BH}$ in a

Run	Resolution (pc)	BH mass ($\mathrm{M_\odot}$)	Accretion radius (pc)	ϵ
low_R	0.40	20	1.6	Slim
med_R	0.10	20	0.4	Slim
low_G	0.16	100	< 1.6	Slim
high_G	0.02	100	< 0.2	Slim
low_G_0.1	0.16	100	< 1.6	0.1
high_G_0.1	0.02	100	< 0.2	0.1

Table 1. Settings of our simulation suite. The second column reports the gravitational resolution (for _G runs) and the highest refinement level resolution (for _R runs). The fourth column is the accretion radius, which is fixed to 4 cells for _R runs and depends on the smoothing length for _G runs. The last column indicates the type of accretion recipe used.

short time, we perform two GIZMO runs setting the radiative efficiency to its custom value, $\epsilon = 0.1$ (low_G_0.1 and high_G_0.1 runs). We report the details of our six simulations in table 1.

3 RESULTS

Figure 3 shows the comparison between the low resolution GIZMO runs with (low_G) and without (low_G_0.1) the slim disc implementation. All the other simulation parameters are the same in the two simulations. It is immediately clear from the comparison that whenever a BH undergoes an intense accretion episode, the large feedback energy available in the radiatively efficient low_G_0.1 case evacuates the BH surroundings, efficiently limiting further BH growth. In the low_G case, on the contrary, even accretion rates significantly higher than $\dot{M}_\mathrm{E}$ result in moderate luminosities that do not impact on the densest gas clumps, and therefore BHs can grow considerably faster. As an example, in the low_G run the most mass growing BH (that will be referred to as $\mathrm{BH_{top}}$ in all runs hereafter, red line in the bottom–left panel of figure 3) reaches a mass larger by up to 2 order of magnitudes compared to the corresponding $\mathrm{BH_{top}}$ in the radiative efficient case at the end of the simulation (red line in the top–left panel panel). The low radiative efficiency of slim discs has then a double effect: first, for any given accretion rate BHs grow faster simply because less mass is lost as radiation (the "$(1-\epsilon)$-effect"); second, the reduced radiative efficiency results in a reduced feedback on the accreting gas, and larger accretion rates are therefore possible (the "$\dot{M}$-effect").

In order to assess how numerical resolution affects our results, we analyse the two high resolution GIZMO runs (high_G and high_G_0.1), and compare the outputs to the low resolution cases discussed above. Figure 4 shows the accretion history of BHs (left panels) and the effect the accretion feedback has on the gas (right panels). Because of the higher resolution we can now resolve a smaller accretion region around each BH, which has the net effect of reducing the BH mass growth compared to the corresponding low resolution runs. Nevertheless, it is apparent how, also in these high resolution simulations, BH mass growth is strongly suppressed in the radiatively efficient case (top panels). Indeed, for $\epsilon = 0.1$, $\mathrm{BH_{top}}$ increases its mass by only $\simeq 50\%$ of its initial value. We stress again that the different radiative efficiency is only marginally responsible of the different accreted mass in the two cases. As clearly shown in the right panels of Figure 4, the largest effect is played by the accretion feedback that, in the standard high-efficiency case, evacuates the region closer to the BHs, hence inhibiting further gas accretion.

The implementation of a physically motivated radiative inefficient accretion mode is then a necessary condition for a fast, highly super–Eddington growth of BHs in our simulations, but, as we will show next, is not sufficient. In the following we will focus only on runs including the slim disc prescription, in order to link episodes of super–Eddington growth with the physical state of the BHs and of the nuclear disc, with the ultimate aim of understanding the processes that can possibly lead to high accretion rates.

Figure 5 shows the results of the highest resolution RAMSES run med_R. The upper left panel reports the mass evolution of the 20 BHs as a function of time. As for the low_G simulation discussed above, the implementation of the slim disc efficiency prescription allows $\mathrm{BH_{top}}$ (shown as a red line) to grow within 3 Myr by up to ~ 3 orders of magnitude in mass. Note that $\mathrm{BH_{top}}$ is not necessary the earliest growing BH of the cluster.

The upper right panel of figure 5 focuses on $\mathrm{BH_{top}}$ alone, showing the time evolution of the accretion rate, and the corresponding distance from the gas clump the BH bounds to during the peak of its mass growth. The accreting clump forms out of a spiral stream developing in the cooling disc, and can not be clearly identified as a bound structure before $t \approx 1$ Myr, as shown in the middle left panel. $\mathrm{BH_{top}}$ passes a first time through the overdense stream (middle right panel), and experiences a short $\lesssim 0.1$ Myr super–Eddington accretion episode, but the radial component of its velocity quickly is large enough to displace it from the overdensity (as observed in the $\dot{m}_\mathrm{BH}$ plot, upper right panel). As the clump grows in mass (up to a maximum of $\sim 3 \times 10^4\,\mathrm{M_\odot}$ in gas), the $\mathrm{BH_{top}}$ feels its gravitational attraction, and is eventually captured by the clump. At this time $\mathrm{BH_{top}}$ undergoes a longer (~ 0.5 Myr) intense super–Eddington accretion phase. Being the initially small BH surrounded by an overwhelmingly large and cold gas cloud, the BH accretes at the maximum rate allowed by the code (i.e. $500 \times \dot{M}_\mathrm{E}$) until almost all gas is turned into stars. At this point $\mathrm{BH_{top}}$ (already grown by 3 order of magnitudes in mass), together with stars exploding as SNae, can evacuate the residual gas condensation (lower right panel). Note that BHs (including $\mathrm{BH_{top}}$) accrete most of their mass from, essentially, a single

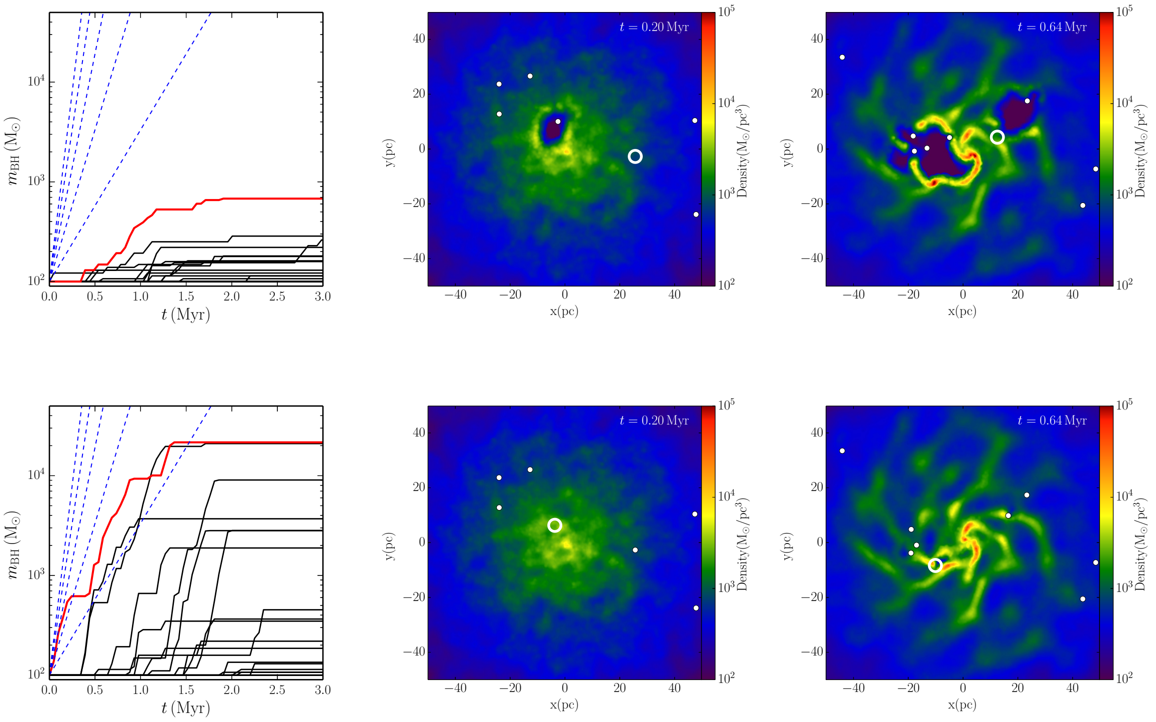

Figure 3. Left panels: BH masses as a function of time for runs low_G_0.1 (top panel) and low_G (bottom panel). The red lines correspond to the most massive BHs (BH$_{\rm top}$) at the end of the runs, while the blue dashed lines trace accretion histories at fixed Eddington ratios of 500, 400, 300, 200 and 100, respectively. Central and right panels: gas density maps for the two runs at $t = 0.2$ and 0.64 Myr, respectively. The white dots mark the positions of the BHs.

dense clump they randomly come across during the dynamical evolution of the system.

It is important to realise that the gravitational capture of a BH by a dense gas clump is intrinsically stochastic, as clumps form in the disc via gravitational instabilities of cooling gas independently of the presence of seed holes. While the BH–capture process is common in all the simulations we ran, the number and mass distributions of gas clumps and, consequently, the fraction of BHs that bind to them, in fact depend upon the spatial and mass resolution we achieve. Figure 6 shows a comparison between runs with different spatial resolution. Among the runs including the slim disc implementation, only run med_R (already shown in Figure 5) is left out of the direct comparison.

A first clear difference is observable at early times. The runs with lower resolution show a faster initial growth of each individual BH, and the number of growing BHs right after the beginning of the runs ($t \lesssim 0.5$ Myr) also increases with decreasing resolution. These trends are caused by the larger accretion radius implemented in the lower resolution runs. In these simulations the BHs can start accreting well before the disc develops any significant overdensity. For this reason the feedback of the early BH accretion onto the gas is more efficient, as a larger energy is injected in a lower density medium. As the resolution increases and the accretion radius can be decreased, fewer BHs have an early start, as in the med_R run (upper left panel of Figure 5) and, more evidently, in the high_G run (lower left panel of Figure 6).

The high_G run, thanks to the exquisite mass and spatial resolutions achieved, shows a richness of structures observable directly in the density map (see the lower right panel of Figure 6 in particular), in which the formation of dense clumps as well as the feedback exerted by the ongoing SF are clearly visible. The gas particles tracing the gas evolution allow us to follow the formation of the massive clump from which BH$_{\rm top}$ gains its mass. Figure 8 reports two different projections of BH$_{\rm top}$ orbit along with the trajectories of 50 gas particles randomly extracted from those forming the massive clump BH$_{\rm top}$ binds to and accretes from. The clump formation clearly proceeds out of a gas gravitational instability within the dense disc, and starts interacting with BH$_{\rm top}$ only when their orbits intersect. Strong gravitational perturbations to the BH orbit are clearly seen when the two systems bind gravitationally. The BH growth then exerts a feedback onto the gas particles, that, together with stars exploding as SNae, results in a partial ejection of particles from the BH neighbourhoods and out of the disc plane (as clearly see in Figure 8 lower panel).

We finally analyse the physical properties of clumps in the high_G run using the public tool SKID[2], and report our results for both the disc-like clumps and the more extended gravitationally bound streams in table 2. Clump properties are evaluated at $t \sim 1.6$ Myr, when $BH_{\rm top}$ is in its main

---

[2] `https://github.com/N-BodyShop/skid`

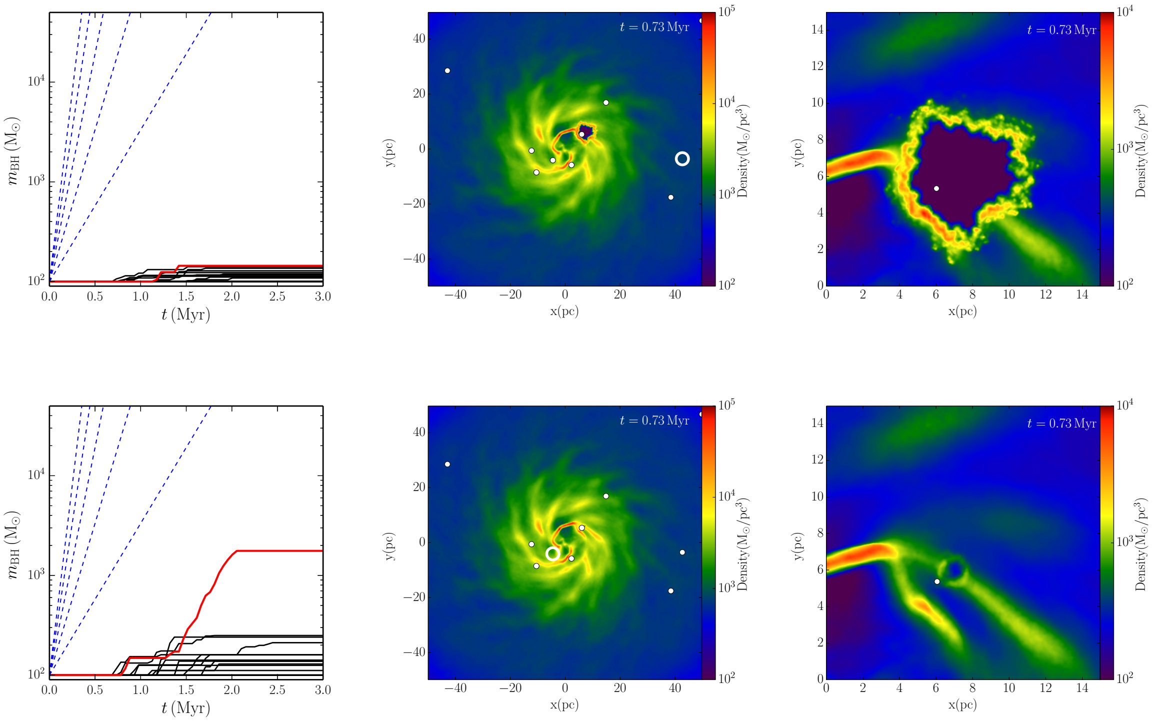

Figure 4. Left panels: BH masses as a function of time for runs high_G_0.1 (top panel) and high_G (bottom panel). The red lines correspond to the most massive BHs ($\rm BH_{top}$) at the end of the runs, while the blue dashed lines trace accretion histories at fixed Eddington ratios of 500, 400, 300, 200 and 100, respectively. Central panels: gas density maps for the two runs at $t = 0.73$ Myr. Right panels: zoom in of a region heated by BH feedback. The white dots mark the positions of the BHs.

accretion phase. Fig. 7 shows the clumps in the disc density map. The reported circular velocity is computed at the half mass radius of the clump. It is worth noticing that typical masses (ranging from few $10^2\,\rm M_\odot$ up to few $10^5\,\rm M_\odot$), as well as the density of our clumps are compatible with the observed properties of local giant molecular clouds (GMCs), though the latter are much less compact (Lombardi, Alves & Lada 2010). The average temperature of the clumps is much higher than the typical temperature of GMCs, mainly because we neglected molecular hydrogen cooling and, at the same time, the metal production in the few Myr of our simulations is not sufficient to significantly alter the gas cooling function. While allowing for a more efficient cooling would initially boost SF in the clumps, further SF could be hampered by the enhanced UV dissociating flux [3].

We conclude by noticing that in our idealised runs the growth of the BHs is finally halted by the star formation–driven gas consumption, and by gas ejection triggered by SNae. However, in a cosmological perspective, the galaxy nucleus would be replenished of gas coming from large scale filaments and/or galaxy mergers. The very short duration of the super–Eddington accretion bursts allows for the growth of stellar mass BHs up to $\gtrsim 10^4\,\rm M_\odot$ or more on a time comparable (or even shorter) than the star–formation timescale.

4 DISCUSSION AND CONCLUSIONS

We presented the results from a suite of numerical high resolution simulations aimed at studying the accretion of stellar mass BHs in nuclear gaseous discs. We implemented a new BH thermal feedback prescription, that takes into account the possible occurrence of radiatively inefficient accretion bursts during which the BHs can actually increase their masses at a significantly super–Eddington pace. We have employed both AMR and Lagrangian mesh-free simulations, achieving comparable results, which strengthens greatly our conclusions.

The set up of our runs is highly idealised, since BHs are supposed to be already in place in a well formed gaseous disc. The latter, at the beginning of the simulation, starts cooling and eventually forms stars. Furthermore, our simulations are evolved in complete isolation, i.e., no gas flows into the nuclear disc from larger scales (e.g., from outer regions of the host galaxy, from cosmological filaments, or through galaxy mergers). As a consequence, every accretion episode halts when star formation and SNa–driven gas consumption have evacuated the central disc regions.

Though the prescriptions adopted in our simulations

---

[3] We can speculate that, if clumps become self-shielded to dissociating radiation and SF continues unimpeded, dense stellar systems could form, which eventually might be prone to dynamical instabilities. A detailed analysis of the formation of dense stellar clusters in our simulations is beyond the scope of the current study.

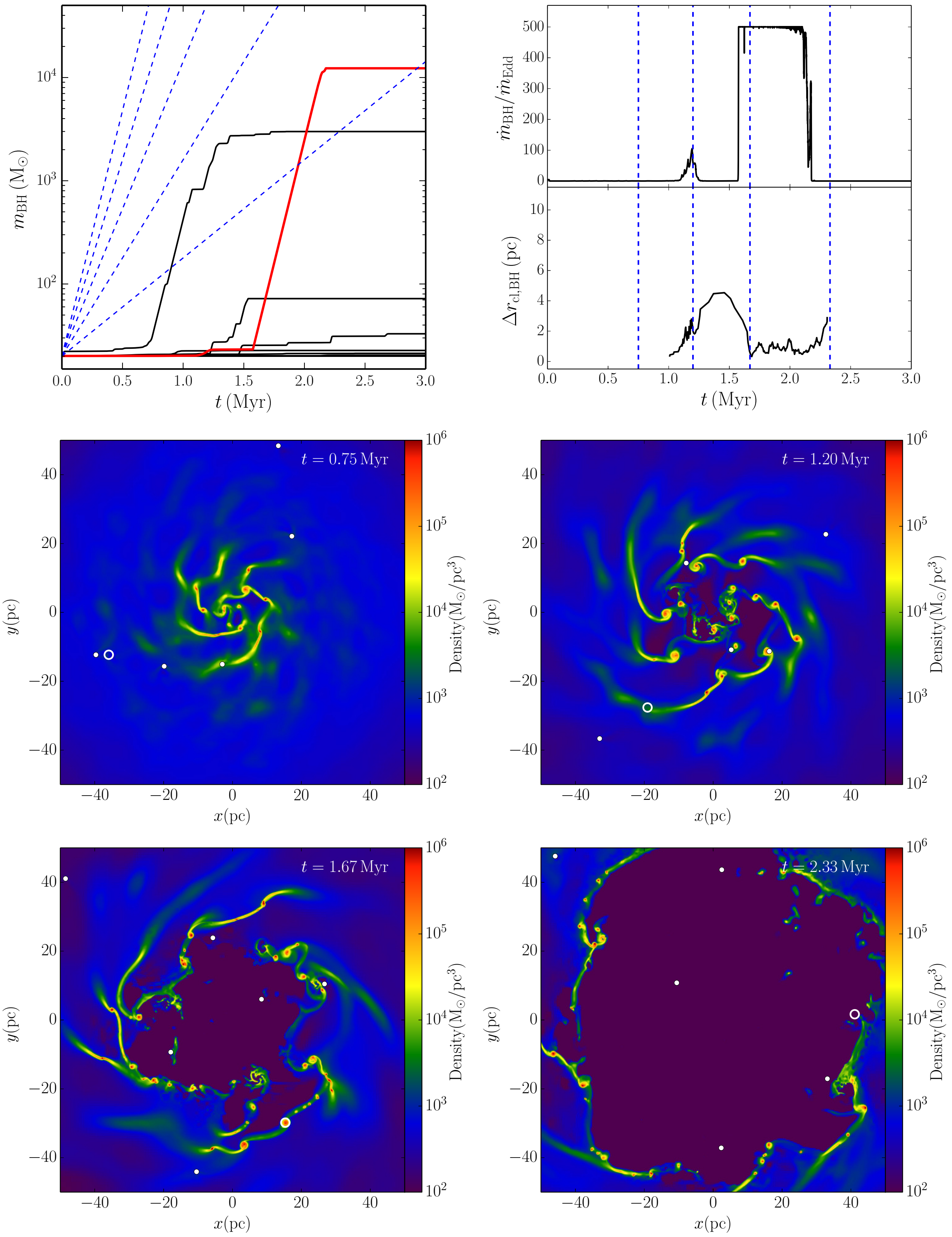

Figure 5. Run med_R. Upper left panel: BH masses vs time for all the 20 BHs. The dashed lines show the slope of accretion episodes at 500, 400, 300, 200 and 100 $\dot{M}_{\rm E}$. Upper right panel: accretion rate for BH$_{\rm top}$, and distance from the clump BH$_{\rm top}$ bounds to during the peak of its mass growth. Middle and lower panels show the density in the equatorial disc plane of the gas at $t = 0.75$, 1.2, 1.67 and 2.33 Myr (corresponding to the times highlighted by the dotted lines in the upper upper right panel). The BH$_{\rm top}$ is reported as large white ring, while the other BHs are shown as smaller white dots.

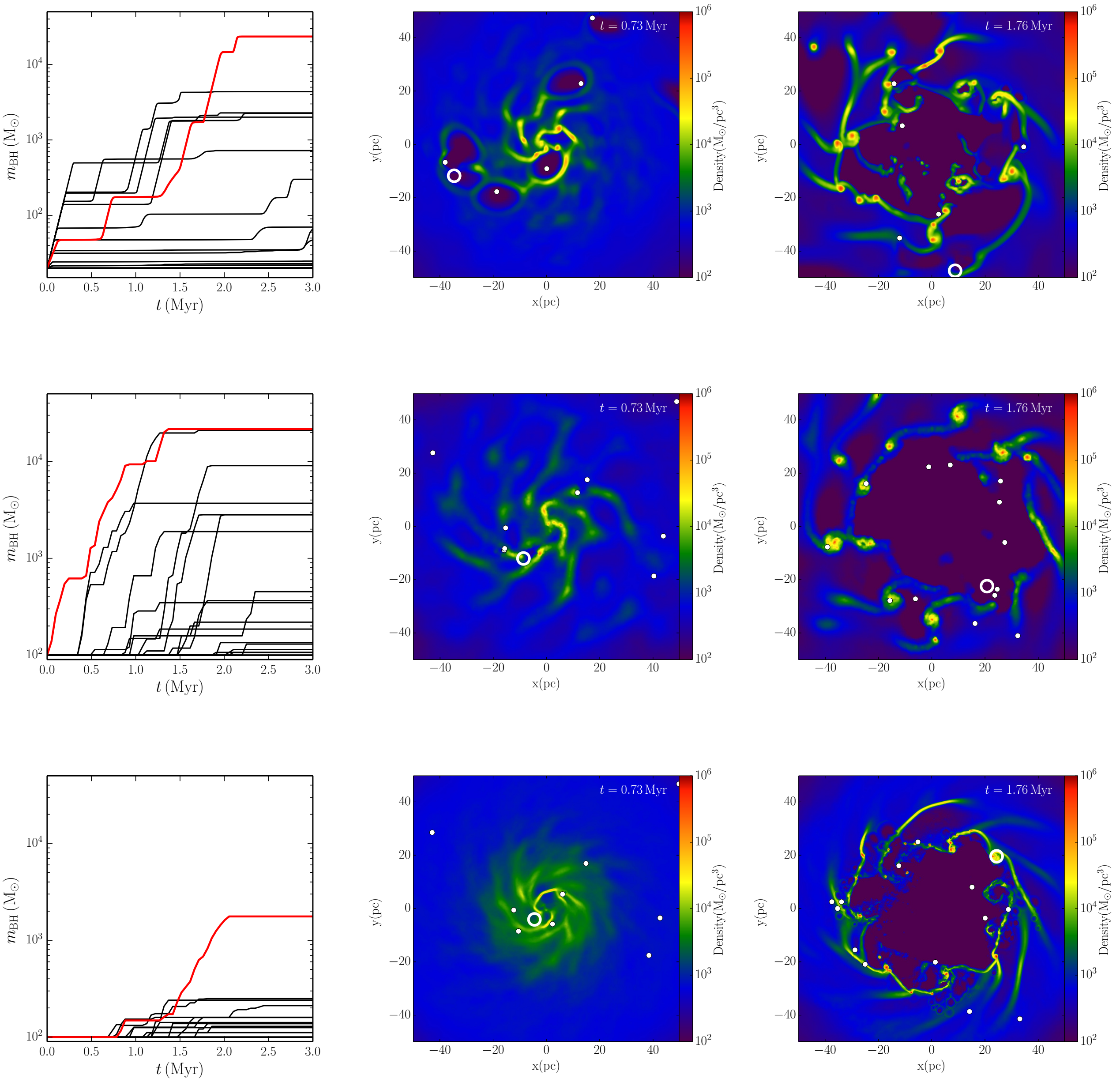

Figure 6. Upper panels: Mass as a function of time of the BHs (left panel), gas density at $t = 0.73$ Myr (central panel) and $t = 1.76$ Myr for the low_R run. The positions of the BHs are shown as white dots. The growth of BH$_{\rm top}$ is highlighted in the left panel with a red line, and its position in the central and right panels is marked with a large white ring. Middle and lower panels, the same as the upper panels for run low_G and high_G, respectively.

regarding star formation and SNa feedback could, in principle, affect the growth of BHs, they are conservative for what concerns gas accretion onto the BHs. First, the star formation rate in a sphere of radius $\simeq 1$ pc (corresponding to the average clump radius) around BH$_{\rm top}$ is $\simeq 0.1\,{\rm M}_\odot$/yr, much larger than the average BH accretion rate ($\simeq 10^{-3}\,{\rm M}_\odot$/yr). Hence, we can be confident that in our simulations the gas is mostly consumed by star formation rather than by BH accretion. To further prove the point we run a low resolution GIZMO simulation in which we increase the star formation efficiency to its maximum value, and find that even for the resulting extremely high star formation rate ($\sim 40\,{\rm M}_\odot$/yr) the BH accretion history is not significantly modified. Second, our assumed timescale for SNa explosions (1 Myr) is shorter than the typical lifetime of low metallicity stars in the mass range $8 - 40\,{\rm M}_\odot$ ($\gtrsim 4$ Myr; Hurley, Pols & Tout (2000)). Our resulting SNa feedback is then already very highly efficient[4]. In this context, we find that SNe produce

---

[4] We run a further GIZMO low resolution simulation and checked

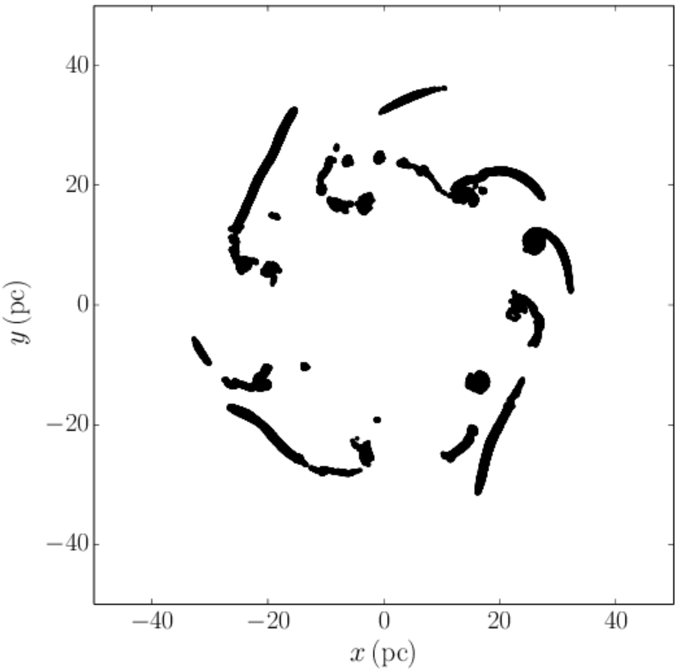

Figure 7. Map of clumps in the high_G run at $t \sim 1.6$ Myr. The clumps can be grouped in two different classes: CL exhibit a disc-like shape and BS are extended bound streams.

M^a	$R^b_{1/2}$	$v^c_{1/2}$	T^d	ρ^e
Bound clumps				
8.70×10^2	0.21	2.98	8.77×10^3	6.77×10^5
1.80×10^3	0.31	3.51	8.89×10^3	7.06×10^5
2.53×10^3	0.24	4.77	8.65×10^3	7.73×10^5
1.68×10^4	0.16	14.96	9.94×10^4	2.26×10^7
3.21×10^4	0.28	15.85	2.91×10^4	5.94×10^6
5.58×10^4	2.48	6.95	1.64×10^4	2.71×10^6
6.75×10^4	0.27	23.22	5.78×10^4	1.31×10^7
8.05×10^4	0.48	18.98	3.47×10^4	7.04×10^6
8.87×10^4	0.48	20.04	2.95×10^4	5.73×10^6
9.38×10^4	0.41	22.27	4.12×10^4	8.40×10^6
1.24×10^5	0.91	17.07	2.23×10^4	4.09×10^6
1.60×10^5	0.70	22.14	3.96×10^4	8.18×10^6
2.14×10^5	1.66	16.64	1.55×10^4	2.51×10^6
Bound streams				
1.13×10^4	0.89	5.24	8.24×10^3	5.45×10^5
3.51×10^4	1.59	6.89	8.43×10^3	5.63×10^5
8.28×10^4	2.55	8.36	2.28×10^4	2.48×10^6
1.70×10^5	3.48	10.24	9.31×10^3	8.05×10^5
1.73×10^5	1.85	14.16	2.41×10^4	4.60×10^6
2.15×10^5	1.16	20.02	3.60×10^4	7.54×10^6
3.25×10^5	7.00	9.99	1.92×10^4	3.30×10^6
3.90×10^5	5.36	12.51	1.20×10^4	1.53×10^6
4.83×10^5	2.91	18.91	1.90×10^4	3.29×10^6

Table 2. Properties of the clumps in the high_G simulation at $t \sim 1.6$ Myr. The bound structures with an almost spherical shape have been classified as bound clumps, while the most asymmetric ones have been classified as bound streams. ^{a} : clump mass in solar masses. ^{b} : clump half mass radius in pc. ^{c} : circular velocity at half mass radius in km/s. ^{d} : gas mean temperatures in K. ^{e} : gas mean number density in cm^{-3}.

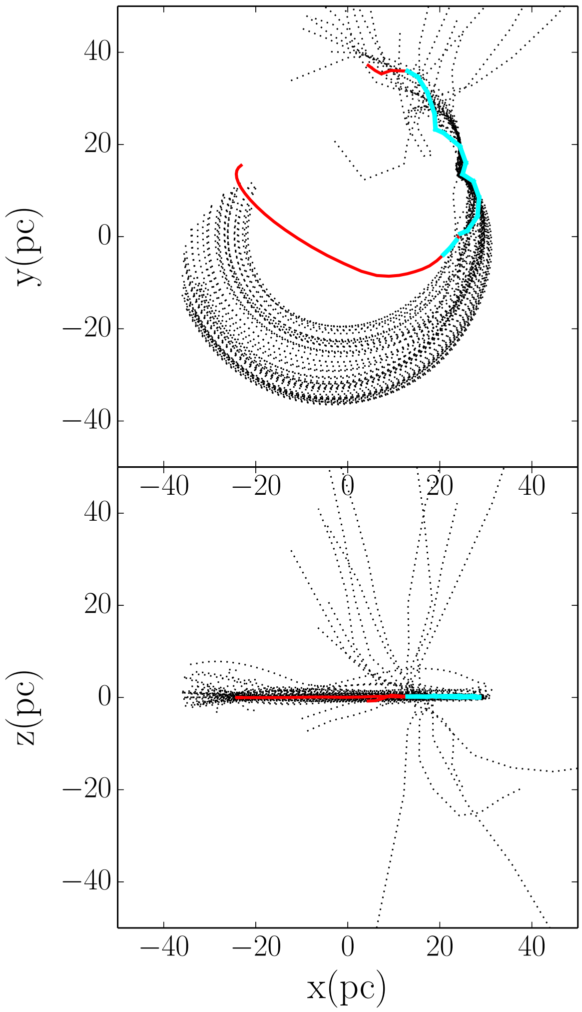

Figure 8. Solid red lines show the face–on (top panel) and edge–on (bottom panel) projections of the trajectory of $\mathrm{BH_{top}}$ in run high_G. Black dotted lines trace the orbits of a sample of the gas particles forming the gas clump $\mathrm{BH_{top}}$ binds to. The accretion burst due to the $\mathrm{BH_{top}}$–clump interaction is highlighted in cyan.

a high velocity wind ($v_{\rm ej} \lesssim 3000\,{\rm km/s}$), which can expand up to 5 kpc above the disc plane. In principle such gas could form a galactic fountain falling back on to the disc, allowing for a new phase of super-critical accretion.

The cooling function employed in our simulations does not take into account molecular hydrogen, which could induce stronger fragmentation in the disc, and enhance SF. This could, in principle, limit the accretion on to the BHs. However, the presence of radiative feedback from stars (not considered in this study) would act in the opposite direction, dissociating molecular hydrogen and thus allowing for higher inflow rates towards the BHs. Both the modelling of the large scale galactic potential (essential to assess the fate of the SNa driven wind) and the effect of molecular hydrogen cooling (and related stellar dissociating flux) are beyond the scope of our study.

Regardless the spatial/mass resolution and the kind of hydrodynamical code used, a coherent picture emerges. If BHs have to grow by 2-3 order of magnitudes in mass, radiative inefficient accretion is a necessary condition, but not a sufficient one. BHs must find themselves embedded in gas structures that need to be: (i) massive enough to provide the gas reservoir, and (ii) dense enough to survive feedback. This may occur when the cooling gas fragments in clumps, and some of the BHs bind to them. Such process allows some of the BHs to reach masses as high as 10^3 - $10^4\,\mathrm{M_\odot}$ on Myr timescales, making them viable candidates as seeds of the supermassive variety of BHs powering high redshift quasars.

Mass accretion onto the BHs depends upon the number, mass and density of the clumps forming in the disc. We showed that these parameters are affected by the numerical resolution achieved in the different runs and, as discussed in Section 3, different resolutions result in different BH accretion histories. We are unable to describe gas dynamics down to the accretion disc scales, even at the highest spatial resolution reached, and this limits our ability to achieve firm estimates of accretion rate and mass growth of the BHs. Yet, the dynamics of gas leading to the formation of dense clouds we observe in all our runs is strongly independent of sub–grid recipes. The gas within the accretion radius of BHs is far from being rotationally supported. Since the relative gas-BH velocity becomes negligible after the capture process, the gas in fact experiences almost radial inflow toward the BHs. Our estimate of the accretion rate is of the same order of the Bondi accretion rate given the temperature and density of the medium surrounding the BHs.

Therefore, despite our accretion histories are not accurate enough from a quantitative point of view, we can be confident about the reliability of the BH-clumps-capture process we observe. Our work should be considered as a *proof of concept*, robust enough to understand under which conditions and through which processes a cluster of stellar mass BHs can actually experience episodes of super–Eddington growth, and what are the effects on the environment.

We conclude that a radiatively inefficient accretion, together with the aforementioned BH-clumps-capture process, can result in mass growths 10-100 times larger than in the radiatively efficient case, making this mechanism a viable candidate to grow massive BH seeds from stellar mass BHs.

The process we studied can result in a prolonged super–Eddington accretion phase only as long as the masses of the clumps are comparable or larger than the masses of the accreting BHs. While the gravitational capture itself easily binds small clumps to comparatively massive BHs, the available gas reservoir is not sufficient for significant BH growth. Moreover, even feedback from radiative inefficient accretion severely affects such small clumps.

Other feedback processes, e.g., momentum-driven feedback, might be important, and will be explored in the future. If, however, such processes turn out to be inefficient, this could naturally allow the galaxy to remain highly star–forming despite the fast growth of the MBH, perhaps explaining the new puzzling observation of a high-z star forming galaxy hosting an SMBH well overweight for its stellar mass (Trakhtenbrot et al. 2015).

that only an unrealistic “maximally efficient” SNa feedback (in which stars explode as soon as they form) is able to evacuate the gas from the centre before the stellar-size BHs can start to accrete.

In addition, as soon as a BH becomes significantly heavier than typical gas clumps, it starts migrating toward the centre of the disc via dynamical friction. This process will naturally bring the most massive BH (the one that by chance had the largest mass growth, i.e., $\mathrm{BH_{top}}$) to the centre of the host galaxy, where MBHs are commonly observed. At this point, however, further clumps forming in the disc no longer interact with the central BH.

In order for the large nuclear gas reservoir assumed in our initial conditions to be present in the galactic nucleus disc angular momentum needs to be removed well before gas turns into stars, so that inflowing material can be, at least partially, accreted by the central BH. This is of course the longstanding fuelling problem of MBH debated in the community (e.g., Hicks et al. 2013, and references therein), and its discussion is beyond the goal of the present study. Paper II will deal with the nuclear properties and gas inflows in a galaxy at $z \sim 6-10$ which by $z = 3$ will have a mass comparable to that of the observed clumpy star forming discs, thus allowing to place our model more properly in the context of galaxy formation and test its assumptions and outcomes. Preliminary analysis shows that the mass enclosed within a hundred pc scale fluctuates between a few times $10^7\mathrm{M_\odot}$ and just above $10^8\mathrm{M_\odot}$. So, as mentioned in Section 2, our initial conditions seem to be well motivated.

We finally note that, whenever inflowing gas refills the circum–nuclear disc, the whole process we simulated is rejuvenated: a new massive BH seed will be formed, sinking to the centre of the galaxy and eventually forming an intermediate massive black hole binary bound to coalesce owing to gravitational radiation losses. This kind of systems may be a perfect target for the planned eLISA observatory (Amaro-Seoane et al. 2013).

5 ACKNOWLEDGEMENTS

We thank the anonymous referee for his/her comments and suggestions that helped us to improve the quality of the paper. AL, FH and MD acknowledge financial support from the Italian MIUR, through PRIN 2010-2011. Simulations were run on the EURORA cluster at CINECA and on the Lucia cluster at DiSAT, University of Insubria. P.M. acknowledges support by the NSF through grant AST-1229745 and NASA through grant NNX12AF87G.

REFERENCES

Abramowicz M. A., Czerny B., Lasota J. P., Szuszkiewicz E., 1988, ApJ, 332, 646

Alvarez M. A., Wise J. H., Abel T., 2009, ApJ, 701, L133

Amaro-Seoane P. et al., 2013, GW Notes, Vol. 6, p. 4-110, 6, 4

Begelman M. C., 2010, MNRAS, 402, 673

Begelman M. C., Rossi E. M., Armitage P. J., 2008, MNRAS, 387, 1649

Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289

Bleuler A., Teyssier R., 2014, MNRAS, 445, 4015

Booth C. M., Schaye J., 2009, MNRAS, 398, 53

Bromm V., Loeb A., 2003, ApJ, 596, 34

Chapon D., Mayer L., Teyssier R., 2013, MNRAS, 429, 3114

Chevalier R. A., 1974, ApJ, 188, 501

Dotan C., Rossi E. M., Shaviv N. J., 2011, MNRAS, 417, 3035

Dubois Y., Volonteri M., Silk J., Devriendt J., Slyz A., 2014, MNRAS, 440, 2333

Fan X. et al., 2006, AJ, 131, 1203

Ferland G. J. et al., 2013, Rev. Mexicana Astron. Astrofis., 49, 137

Ferrara A., Haardt F., Salvaterra R., 2013, MNRAS, 434, 2600

Ghisellini G., Celotti A., Tavecchio F., Haardt F., Sbarrato T., 2014, MNRAS, 438, 2694

Haiman Z., Loeb A., 2001, ApJ, 552, 459

Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, ApJ, 591, 288

Hicks E. K. S., Davies R. I., Maciejewski W., Emsellem E., Malkan M. A., Dumas G., Müller-Sánchez F., Rivers A., 2013, ApJ, 768, 107

Hopkins P. F., 2014, ArXiv e-prints, 1409.7395

Hurley J. R., Pols O. R., Tout C. A., 2000, MNRAS, 315, 543

Kennicutt, Jr. R. C., 1998, ApJ, 498, 541

Kim J.-h. et al., 2014, ApJS, 210, 14

Koushiappas S. M., Bullock J. S., Dekel A., 2004, MNRAS, 354, 292

Latif M. A., Bovino S., Grassi T., Schleicher D. R. G., Spaans M., 2015, MNRAS, 446, 3163

Loeb A., Rasio F. A., 1994, ApJ, 432, 52

Lombardi M., Alves J., Lada C. J., 2010, A&A, 519, L7

Lupi A., Haardt F., Dotti M., 2015, MNRAS, 446, 1765

Madau P., Haardt F., Dotti M., 2014, ApJ, 784, L38

Madau P., Rees M. J., 2001, ApJ, 551, L27

Madau P., Rees M. J., Volonteri M., Haardt F., Oh S. P., 2004, ApJ, 604, 484

Mayer L., Fiacconi D., Bonoli S., Quinn T., Roskar R., Shen S., Wadsley J., 2014, ArXiv e-prints, 1411.5683

Mayer L., Kazantzidis S., Escala A., Callegari S., 2010, Nature, 466, 1082

Medling A. M. et al., 2014, ApJ, 784, 70

Milosavljević M., Bromm V., Couch S. M., Oh S. P., 2009, ApJ, 698, 766

Mortlock D. J. et al., 2011, Nature, 474, 616

Russell D. M., Gallo E., Fender R. P., 2013, MNRAS, 431, 405

Sądowski A., Narayan R., McKinney J. C., Tchekhovskoy A., 2014, MNRAS, 439, 503

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Spaans M., Silk J., 2006, ApJ, 652, 902

Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074

Tanaka T., Haiman Z., 2009, ApJ, 696, 1798

Teyssier R., 2002, A&A, 385, 337

Teyssier R., Pontzen A., Dubois Y., Read J. I., 2013, MNRAS, 429, 3068

The Enzo Collaboration, Bryan G. L. et al., 2014, ApJS, 211, 19

Trakhtenbrot B. et al., 2015, Science, 349, 168

Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559

Volonteri M., Silk J., Dubus G., 2015, ApJ, 804, 148

Wise J. H., Turk M. J., Abel T., 2008, ApJ, 682, 745

Accepted for publication in ApJ
Preprint typeset using LaTeX style emulateapj v. 01/23/15

COLD DARK MATTER SUBSTRUCTURES IN EARLY-TYPE GALAXY HALOS

Davide Fiacconi[1], Piero Madau[1,2,3], Doug Potter[1], and Joachim Stadel[1]

Accepted for publication in ApJ

ABSTRACT

We present initial results from the "Ponos" zoom-in numerical simulations of dark matter substructures in massive ellipticals. Two very highly resolved dark matter halos with $M_{\rm vir} = 1.2 \times 10^{13}\,{\rm M}_\odot$ and $M_{\rm vir} = 6.5 \times 10^{12}\,{\rm M}_\odot$ and different ("violent" vs. "quiescent") assembly histories have been simulated down to $z = 0$ in a ΛCDM cosmology with a total of 921,651,914 and 408,377,544 particles, respectively. Within the virial radius, the total mass fraction in self-bound $M_{\rm sub} > 10^6\,{\rm M}_\odot$ subhalos at the present epoch is 15% for the violent host and 16.5% for the quiescent one. At $z = 0.7$, these fractions increase to 19 and 33%, respectively, as more recently accreted satellites are less prone to tidal destruction. In projection, the average fraction of surface mass density in substructure at a distance of $R/R_{\rm vir} = 0.02$ ($\sim$ 5–10 kpc) from the two halo centers ranges from 0.6% to $\gtrsim$ 2%, significantly higher than measured in simulations of Milky Way-sized halos. The contribution of subhalos with $M_{\rm sub} < 10^9\,{\rm M}_\odot$ to the projected mass fraction is between one fifth and one third of the total, with the smallest share found in the quiescent host. We assess the impact of baryonic effects via twin, lower-resolution hydrodynamical simulations that include metallicity-dependent gas cooling, star formation, and a delayed-radiative-cooling scheme for supernova feedback. Baryonic contraction produces a super-isothermal total density profile and increases the number of massive subhalos in the inner regions of the main host. The host density profiles and projected subhalo mass fractions appear to be broadly consistent with observations of gravitational lenses.

Keywords: cosmology: theory – dark matter – galaxies: halos – gravitational lensing: strong – methods: numerical

1. INTRODUCTION

Precision measurements across cosmic time have made the cold dark matter plus cosmological constant (ΛCDM) paradigm one of the pillars of our current understanding of the origin and evolution of structures in the universe. This model has proven to be remarkably successful at matching observations from scales that span the horizon length (e.g., Planck Collaboration et al. 2015) all the way down to the scales probed by the Lyman-α forest (e.g., Viel et al. 2013). A vast zoo of non-baryonic dark matter candidates has been proposed over the last three decades to reproduce the wealth of cosmological/astrophysical data (e.g., Feng 2010).

The standard theory of structure formation requires dark matter to be cold, i.e. made of particles that become non-relativistic well before the matter domination era, and therefore clump on all scales. It is precisely on the smallest subgalactic scales that there have been persistent observational challenges to the cold, collisionless dark matter expectations. These "small-scale controversies", predominantly found in the abundances and density profiles of dark matter-dominated dwarfs in the local universe, may simply stem from a poor understanding of the baryonic processes involved in galaxy formation (e.g. Pontzen & Governato 2014; Madau et al. 2014; Weinberg et al. 2015). They may alternatively indicate the need for more complex physics in the dark sector, and many modifications of the properties of the dark matter particle have been proposed to suppress small-scale power and alleviate some of these problems, including warm dark matter (WDM; Bode et al. 2001), self-interacting dark matter (Spergel & Steinhardt 2000), fuzzy dark matter (Hu et al. 2000), and superWIMPS (Cembranos et al. 2005).

There is, however, no consensus on how serious these problems really are for ΛCDM, and detailed testing of the standard paradigm on small scales remains one of the most pressing issues in cosmology. Numerical simulations in ΛCDM have shown a rich spectrum of substructures in galaxy halos, the fossil remnants of a hierarchical merging process that is never complete (Moore et al. 1999; Klypin et al. 1999; Diemand et al. 2007, 2008; Springel et al. 2008; Stadel et al. 2009). Metcalf & Madau (2001) first showed that self-bound subhalos would have a readily detectable gravitational lensing effect if they were as numerous as predicted in ΛCDM. In particular, small fluctuations in the galaxy-scale lensing potential caused by substructures should result in measurable perturbations in the relative magnifications of quadruply-lensed quasar images. Discrepancies between the observed flux ratios and those predicted by a smooth lens model ("flux ratio anomalies") have indeed been found to be common in quasar lenses (e.g., Mao & Schneider 1998; Chiba 2002; Metcalf & Zhao 2002), and can be explained if a large fraction ($\sim$ 2%) of the projected mass at the Einstein radius is in the form of local substructures (Dalal & Kochanek 2002).

To date, there is still no general agreement on whether semi-analytic models or N-body simulations in ΛCDM predict enough substructures to explain the frequency

---

fiacconi@physik.uzh.ch
[1] Center for Theoretical Astrophysics and Cosmology, Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
[2] Institute for Astronomy, ETH Zurich, CH-8093 Zürich, Switzerland
[3] Department of Astronomy & Astrophysics, University of California, 1156 High Street, Santa Cruz, CA 95064, USA

of lens anomalies in currently available samples (e.g., Bradač et al. 2004; Macciò et al. 2006; Amara et al. 2006; Xu et al. 2009; Chen et al. 2011; Metcalf & Amara 2012; Xu et al. 2015). The discrepancy largely arises because of the small number of subhalos anticipated to survive near the radii where images form (typically around 5–10 kpc in projection). Somewhat surprisingly, most numerical simulation work has focused on present-day Milky Way-sized halos as the lens system. And yet, most gravitational lens galaxies are known to be early-type massive galaxies in relatively low density environments (e.g. Kochanek et al. 2000), while the average deflector redshift of the multiply-imaged quasars in the Cosmic Lens All-Sky Survey (CLASS) is $\langle z_{\rm lens}\rangle \approx 0.6$ (Browne et al. 2003). The abundance of substructures in a host is set by the competition between tidal disruption and new accretion. More massive hosts (as well as hosts at higher redshifts) are then expected to be more clumpy because their subhalos have been accreted more recently and managed to survive tidal destruction (Zentner et al. 2005; Klypin et al. 2011).

In order to refine ΛCDM predictions for substructure lensing, we have initiated a program, dubbed "Ponos", of very high resolution N-body and hydrodynamic cosmological simulations of early-type massive galaxies. In this Paper we present initial results on halo substructures from new "zoom-in" simulations of two $\sim 10^{13}\,{\rm M}_\odot$ systems, each having between 155 and 286 million dark matter particles within the virial radius $R_{\rm vir}$ (defined here as the radius enclosing a mean matter density $\Delta\,\rho_c$, where ρ_c is the critical density and Δ is the redshift-dependent virial overdensity, Bryan & Norman 1998). On these halo mass scales, the global fraction of E/S0 galaxies is estimated to be about 50% (Wilman & Erwin 2012). We focus on very high resolution calculations of a limited number of halos to study in detail the relative impact of light and heavy subhalos in strong lensing produced by elliptical galaxies, while we plan to extend this study to a larger halo sample. The two target halos are part of the AGORA Simulation Project[4] (Kim et al. 2014).

2. NUMERICAL SIMULATIONS

The Ponos collisionless simulations have been performed with the Tree/N-body code PKDGRAV3. Gravity was computed using a fast multipole method (Dehnen 2000, 2002) based on a 5th-order reduced expansion for faster and more accurate force calculation, and a multipole based Ewald summation technique for periodic boundary conditions (Stadel et al. 2009; Stadel 2013). An opening angle of $\theta = 0.55$ was adopted for the gravity tree. Time-steps were assigned individually to each particle based on the local dynamical time $\Delta t \leq \xi/\sqrt{G\rho}$, where $\xi = 0.03$ is an accuracy parameter and ρ is the density enclosed within the particle orbit (Zemp et al. 2007). This time-stepping strategy is both faster and more adaptive and accurate than the conventional time-step $\Delta t \leq \xi\sqrt{\epsilon_g/|\mathbf{a}|}$, where $|\mathbf{a}|$ is the magnitude of the local acceleration and ϵ_g is the gravitational softening.

Initial conditions (ICs) were generated with the MUSIC[5] package (Hahn & Abel 2011), and are based upon a *Wilkinson Microwave Anisotropy Probe* cosmology with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$, $\Omega_b = 0.0455$, $\sigma_8 = 0.807$, $n_s = 0.961$, and $H_0 = 70.2\,{\rm km\ s^{-1}\,Mpc^{-1}}$ (Hinshaw et al. 2013). The two target halos were selected from a low resolution pathfinder simulation of a 85.5 comoving Mpc box, and are characterized by very different assembly histories (Kim et al. 2014). One ("PonosQ", $M_{\rm vir} = 6.5\times10^{12}\,{\rm M}_\odot$) has a relatively *quiescent* early merger history (i.e. few major mergers between $z = 4$ and 2, as defined by Kim et al. 2014), while the other ("PonosV", $M_{\rm vir} = 1.2\times10^{13}\,{\rm M}_\odot$) is characterized by a more *violent* early merger history (i.e. many mergers between $z = 4$ and 2) [6]. A strong isolation criterion was imposed on the quiescent halo, one in which its $3R_{\rm vir}$ radius circle does not intersect the $3R_{\rm vir}$ radius circle of any halo with half or more of its mass at $z = 0$. A relaxed criterion was used for the most massive violent halo: $2R_{\rm vir}$ circle instead of $3R_{\rm vir}$. Higher-resolution simulations were performed on new ICs re-centered on each of the target halos, for a total of six further nested spatial refinements (each by a factor of 2), corresponding to an effective resolution of 8192^3 particles with mass $m_p = 4.2\times10^4\,{\rm M}_\odot$. The final highest-resolution region is the convex hull Lagrangian volume that contains all the particles that fall within $3R_{\rm vir}$ of the target halo at $z = 0$, sufficiently large to include all the structures that merge with the main host or have a significant impact on its evolution. The high resolution regions of PonosV and PonosQ were sampled with 890 and 388 million particles, respectively, and evolved with a force resolution $\epsilon_g = 210$ pc (fixed in physical units after $z = 9$ and equal to 1/50 of the initial interparticle separation) from $z = 100$ to $z = 0$. Contamination within $R_{\rm vir}$ from coarse-level particles was found to be well below 0.1% in both number of particles and mass fraction down to the present epoch. To check for numerical convergence, both halos were re-simulated at a factor two lower spatial resolution. Some basic properties of the simulated target halos are listed in Table 1.

2.1. *Halo finding, merger-trees, and density profiles*

Dark matter field halos and subhalos were identified with the AMIGA HALO FINDER (AHF; Gill et al. 2004; Knollmann & Knebe 2009). AHF employs a recursively refined grid to locate local overdensities in the density field. The identified density peaks are treated as centers of prospective halos, subhalos, etc. An iterative procedure is then used to collect those particles that are bound to those centers while removing those that are unbound. A bound clump of mass m at a distance d from the center of a more massive, $M > m$, halo is a subhalo of M if $d < R+sr$, where r and R are the virial (or tidal) radii of m and M, respectively, and $s = 0.75$ is a superposition parameter. Note that, according to this criterion, clumps whose centers lie just outside R can still be subhalos of M. We have checked the reliability of the identification of subhalos by comparing the AHF results with those produced by Rockstar (Behroozi et al. 2013). We find nice agreement between the two halo finders in terms of subhalo mass functions, positions, and distributions of

---

[4] sites.google.com/site/santacruzcomparisonproject/.

[5] www.phys.ethz.ch/ hahn/MUSIC.

[6] We note that, despite the naming of the simulations, PonosQ has a more active assembly history at low redshift than PonosV, as discussed in Section 2.1.

Table 1
Properties of the simulated elliptical halos.

Name	Merger History ($2 < z < 4$)	z_f	$M_{\rm vir}$ ($\rm M_\odot$)	$R_{\rm vir}$ (kpc)	$V_{\rm host}$ ($\rm km\,s^{-1}$)	r_s (kpc)	c	$N_{\rm vir}$	$N_{\rm sub}$
PonosV	violent	1.10	1.2×10^{13}	600.5	348.2	50.9	11.8	285,558,928	48,681
PonosQ	quiescent	0.74	6.5×10^{12}	489.6	265.3	55.0	9.06	154,489,668	26,521

Note. — Columns 3, 4, 5, 6, 7, 8, 9, and 10 give the formation redshift, present-day virial mass, virial radius, maximum circular velocity, scale radius and concentration parameter of the best-fit Navarro et al. (1997) (NFW) spherically-averaged density profile, the number of dark matter particles and the number of self-bound subhalos within the virial radius of the target halos, respectively.

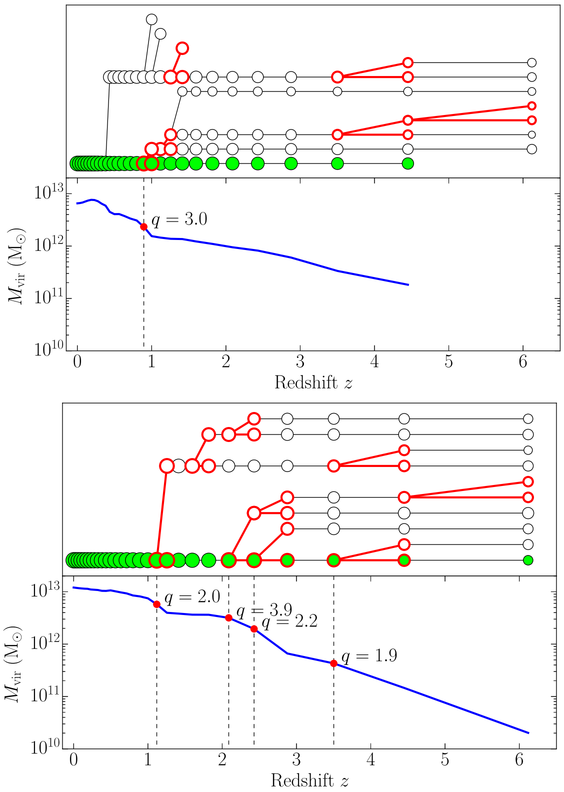

Figure 1. The "merger trees" of our simulated elliptical-sized halos, PonosQ (top) and PonosV (bottom). The main halo at the present day is plotted in the bottom-left corner, and all its progenitors (and also their histories) are depicted backward in time. The size of the symbols scales with halo mass, and the "main branch" is colored in green. Only mergers with mass ratios $q \equiv M/m < 8$ are shown, and all major mergers with mass ratios $q < 4$ are highlighted with the red color. The bottom portion of each panel shows the mass growth history of the main host, with major mergers marked by the vertical dashed lines.

virial (tidal) radii[7].

Figure 1 shows the merger trees of the two main hosts, including only mergers with mass ratios $q \equiv M/m < 8$. We traced subhalos backward and forward in time, matching particles with the same ID inside each system for all snapshots between redshift 10 and 0. The main progenitor of a parent halo in a given snapshot was identified as the halo in the previous snapshot that maximizes the ratio $N_s/\sqrt{Nn}$, where N and n are the number of particles in the parent and progenitor, respectively, and N_s is the number of particles they share (e.g., Fiacconi et al. 2015). As already mentioned above, PonosV and PonosQ have rather different assembly histories. While more massive, PonosV forms earlier (its formation redshift, defined as the redshift at which the halo had already assembled half of its mass, is $z_f = 1.1$) and undergoes four major mergers with $q < 4$ at $z \lesssim 3.5$. PonosQ forms a bit later ($z_f = 0.74$) and has only one major merger (with $q = 3$) at $z \simeq 0.9$. After $z = 3$, PonosV mostly grows within a main dark matter filament that routes the mass accretion toward the halo. On the other hand, PonosQ assembles around $z \sim 1 - 2$ after the merger of a net of short dark matter filaments converging to the position of the halo. The latter grows quickly after $z \sim 1$ and gathers most of the nearby smaller halos, cleaning out the surroundings consistently with the more stringent isolation criterion adopted for the selection.

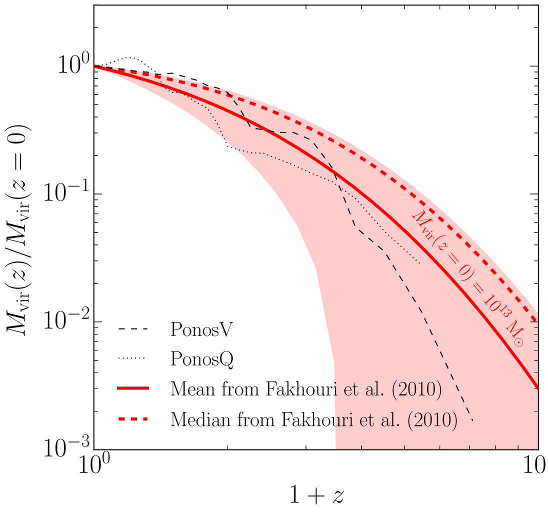

Figure 2. Mass growth history of both PonosQ (thin dotted line) and PonosV (thin dashed line) as a function of redshift z. Red thick lines show the mean (solid) and median (dashed) mass growth of a 10^{13} $\rm M_\odot$ halo at $z = 0$ as determined by Fakhouri et al. (2010). The red shaded region shows the estimate of the standard deviation on the relation. The comparison reveals that both PonosQ and PonosV have a fairly representative (though different) assembly history, as they are both compatible with the mean growth of a 10^{13} $\rm M_\odot$ halo.

Figure 2 compares the mass growth of PonosQ and PonosV with the mean and median growth of a 10^{13} $\rm M_\odot$ halo at $z = 0$. These quantities have been calculated from the z-dependent mass accretion rates as inferred

---

[7] By default, Rockstar does not output the tidal radius of a halo. We extended the code by computing the tidal radius as the distance from the halo center of the farthest bound particle.

from the data of both the Millennium and Millennium-II simulation (see equation (2) of Fakhouri et al. 2010). PonosV follows the average growth after $z \sim 3$, while PonosQ mostly stays slightly below the curve (since it is lighter), but then it gets closer to the fit after $z \sim 0.9$, as it assembles later than PonosV. Although we do not know the dispersion on the fitted relations, we can roughly estimate the standard deviation σ from the relation $|m - \mu|/\sigma \leq \sqrt{3/5}$, where μ and m are the mean and median, respectively (Basu & DasGupta 1997). Figure 2 shows that both PonosQ and PonosV have mass growths compatible with the average at the present-day reference mass 10^{13} $\mathrm{M}_\odot$; this generally qualifies both of them as "representative" halos for their mass scale, though they behave rather differently in detail and that influences the properties of subhalos (see Section 3 and 4).

Figure 3 shows the present-day spherically-averaged density profiles, their logarithmic slope $d \ln \rho / d \ln r$, and the fractional deviations from the best-fit NFW profiles for the two target halos. The NFW formula with scale radii of 51 and 55 kpc and concentration parameters of 11.8 and 9.1 for PonosV and PonosQ, respectively, produces a reasonable approximation to the density profiles down to the convergence radius of each simulation. At all radii between 1 and 500 kpc, deviations from the best-fit NFW matter densities are typically less than 20%.

3. SUBSTRUCTURE ABUNDANCE

We counted all substructures down to a minimum of 20 bound particles, corresponding to a minimum subhalo mass of $M_{\mathrm{res}} = 8.4 \times 10^5$ $\mathrm{M}_\odot$. Our selection results in a sample of 48,681 individual subhalos at $z = 0$ in PonosV, and 26,521 subhalos in PonosQ. In Figure 4 we present the cumulative maximum circular velocity function (normalized to the maximum circular velocity of the host) of the subhalo population of each Ponos host. These are compared with the subhalo velocity functions of the Via Lactea II simulation (Diemand et al. 2008). Over the interval $0.02 < x < 0.08$, all velocity functions can be approximated by the power-law

$$N(> x) = 3.6 \times 10^{-3}\, x^{-3}\, V_{\mathrm{host}}^{1/2}, \tag{1}$$

where $x \equiv V_{\mathrm{sub}}/V_{\mathrm{host}}$ and V_{host} is given in units of $\mathrm{km\,s^{-1}}$ (c.f., Klypin et al. 2011). Residuals from the power-law regression line are typically of order 20%. At a given $V_{\mathrm{sub}}/V_{\mathrm{host}}$, the number $N(> x)$ of subhalos is 56% higher in PonosV compared to Via Lactea, and 86% higher in PonosQ. This is because more massive ellipticals are dynamically younger than Milky Way-sized galaxies, and have more subhalos that manage to survive tidal destruction. Note the relatively large deviations above $x \gtrsim 0.15$ that are present in all three hosts compared to the fitting formula.

Figure 5 shows the cumulative mass fraction in self-bound substructures of PonosQ and PonosV as a function of M_{sub},

$$f_{\mathrm{sub}}(< M_{\mathrm{sub}}) = \frac{1}{M_{\mathrm{vir}}} \int_{M_{\mathrm{res}}}^{M_{\mathrm{sub}}} m\, \frac{dN}{dm}\, dm, \tag{2}$$

at three different redshifts, $z = 0, 0.5, 0.7$. Here, dN/dm is the subhalo mass function. At the present epoch, we measure a total mass fraction $f_{\mathrm{sub}} = 15\%$ in the violent host and $f_{\mathrm{sub}} = 16.5\%$ in the quiescent one. At $z = 0.7$, these fractions increase to 19% and 33%, respectively, as more recently accreted subhalos are more likely to survive tidal stresses. Compared to PonosV, PonosQ forms later, is less concentrated, and undergoes the majority of its mergers at lower redshifts. As a consequence, there is less time for the orbital decay and mass loss of infalling satellites, and its substructure mass fraction is higher (e.g., Chen et al. 2011).

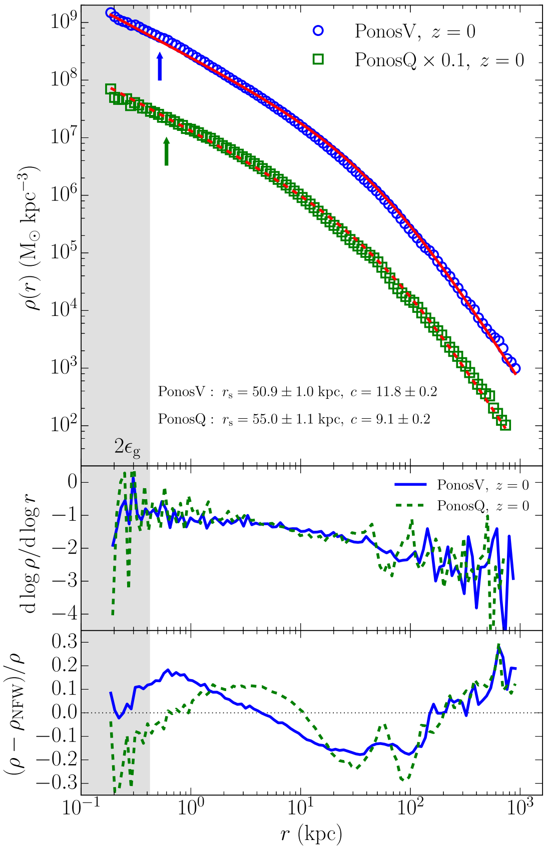

Figure 3. Top: present-day density profiles of PonosV (blue circles) and PonosQ (green squares). The red solid and dashed lines are the best-fit NFW profiles. The density of PonosQ has been rescaled by a factor of 1/10 for clarity. The colored arrows indicates the regions within which numerical convergence is not achieved because of two-body relaxation (Power et al. 2003). Middle: logarithmic slope of the matter density profiles of the two target halos. Bottom: the residuals between the density profile and the best-fit NFW profile, as a function of radius. In all panels, the gray shading indicates the region inside two gravitational softening lengths.

The definition of virial radius, while formally meaningful, is nevertheless rather arbitrary. We have checked that companion systems just outside the virial radius (in the spherical shell $R_{\mathrm{vir}} < r < 2R_{\mathrm{vir}}$) would make a relatively modest contribution to the "subhalo" mass fraction, corresponding to less than a 30% increase in f_{sub}. This is not true, however, for subhalos in the shell $R_{200} < r < R_{\mathrm{vir}}$. To facilitate comparison with previous simulations, we have computed f_{sub}^{200}, the subhalo mass fraction within R_{200}, the radius with mean enclosed overdensity equal to 200 times the critical value[8]. At the present epoch, we derive f_{sub}^{200} =6.9% in PonosV and

---

[8] At $z = 0$, we measure $R_{200} = 444.4$ and 359.6 kpc for PonosV and PonosQ, respectively.

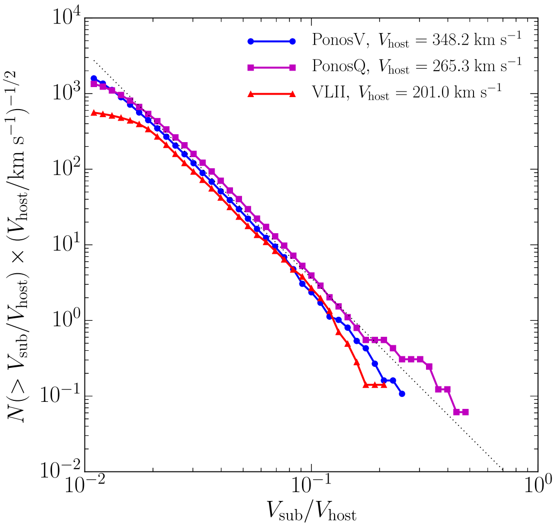

Figure 4. Cumulative subhalo abundance at $z = 0$ as a function of maximum subhalo circular velocity, $V_{\rm sub} \equiv \max\{\sqrt{GM_{\rm sub}(< r)/r}\}$, in units of the maximum circular velocity of the main host, $V_{\rm host}$. We show results for PonosV and PonosQ and, for comparison, for the Via Lactea II simulation. The dotted line shows the best-fit power-law relation of Equation (1).

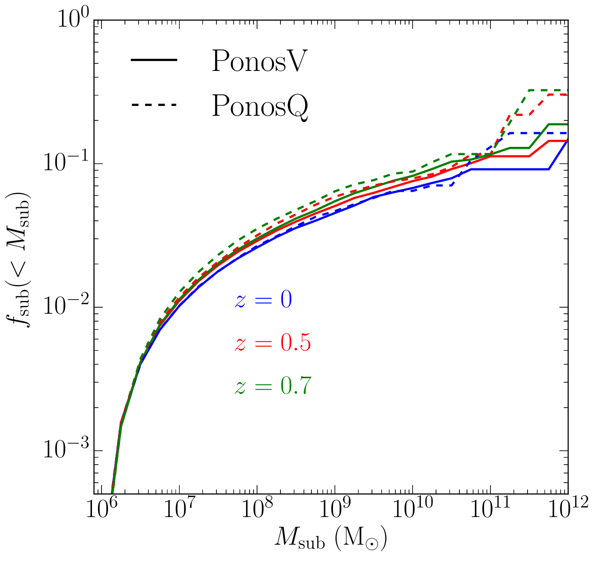

Figure 5. The cumulative substructure mass fraction of PonosQ (dashed lines) and PonosV (solid lines), as a function of subhalo mass $M_{\rm sub}$, at z =0 (blue), 0.5 (red), and 0.7 (green).

$f_{\rm sub}^{200}$ =11.4% in PonosQ. These values are consistent with the mean mass fraction inferred at $z = 0$ in the six Milky Way-sized halos simulated at resolution level 2 as part of the Aquarius project, $\langle f_{\rm sub}^{200} \rangle = (7.2 \pm 2.3)$%, as reported by Xu et al. (2009). However, subhalos within R_{200} account for only 40% (PonosV) and 57% (PonosQ) of the total subhalo mass fraction $f_{\rm sub}$ within $R_{\rm vir}$, the remaining substructure material being located between R_{200} and $R_{\rm vir}$.

According to Xu et al. (2009), the scatter among the six Aquarius halos is much larger than the differences between halos at $z = 0$ and $z = 0.6$. By contrast, the two Ponos hosts have a larger mass fraction in substructure at $z = 0.7$ than at present: $f_{\rm sub}^{200}$ =12.1% in PonosV and $f_{\rm sub}^{200}$ =21.2% in PonosQ. These values are twice as large as measured at $z = 0$. These large variations are associated with the late accretion of a few, relatively massive, $M_{\rm sub} \gtrsim 10^{11}\,{\rm M_\odot}$ satellites. At $z = 0.7$ there are two such systems in the outer halo of PonosV, and four in PonosQ. By $z = 0$, only one of them survives tidal stripping and remains above $10^{11}\,{\rm M_\odot}$ in PonosV, and two in PonosQ. We note, in passing, that late minor mergers with progenitors mass ratios $q < 30$ are fairly typical for descendants of mass $M_{\rm vir} \approx 10^{13}\,{\rm M_\odot}$ today: an analysis of the joint data set from the Millennium and Millennium-II simulations gives rates for such events that exceeds 1 Gyr^{-1} per halo at redshift 1 (Fakhouri et al. 2010).

Over the interval $10^8\,{\rm M_\odot} < M_{\rm sub} < 10^{11}\,{\rm M_\odot}$, our measured subhalo mass function is well approximated by the power law

$$\frac{dN}{d\ln M_{\rm sub}} = N_0 \left(\frac{M_{\rm sub}}{m_0}\right)^{-n}. \tag{3}$$

At $z = 0$, we derive for PonosV a best-fit slope of $n = 0.877$ and an amplitude at the pivot mass, $m_0 = 10^8\,{\rm M_\odot}$, of $N_0 = 847$. At $z = 0.7$, the best-fit parameters are $n = 0.915$ and $N_0 = 868$. A similar relation holds also for PonosQ, with slopes and amplitudes $(n, N_0) = (0.754, 362)$ at $z = 0$, and $(n, N_0) = (0.949, 344)$ at $z = 0.7$. In the redshift interval $0 \leq z \leq 0.7$, the most massive subhalo reaches 8% of the mass of the host. The expected total mass fraction in self-bound substructure below our nominal resolution limit of $M_{\rm res} = 8.4 \times 10^5\,{\rm M_\odot}$ is

$$f_{\rm sub}(< M_{\rm res}) = \frac{N_0 m_0}{M_{\rm vir}(1-n)} \left(\frac{M_{\rm res}}{m_0}\right)^{1-n} \tag{4}$$

(assuming a thermal free-streaming mass limit → 0). The above best-fit power laws yield between one and three percent of the virial mass in unresolved subhalos at $z = 0$, considerably smaller compared to the fractional mass in substructures that are already resolved in our simulations. The substructure mass fraction converges much more slowly at $z = 0.7$, however, as the slope of the subhalo mass function is closer to $n = 1$. Resolved subhalos do not trace the matter distribution of the host: tidal disruptions are most effective in the inner halo, leading to an antibias in the abundance profile of substructures relative to the smooth background (Diemand et al. 2007). At the present epoch, the subhalo mass fraction within the inner 10 kpc drops to 2×10^{-4} and 1.3×10^{-6} in the violent and quiescent host, respectively. We caution, however, about resolution effects (numerical "overmerging") on this quantity. Projecting the subhalos relative to the total mass along the line-of-sight (see next section) will be less strongly influenced by the incompleteness of the inner regions of the host.

4. SUBSTRUCTURE SURFACE MASS DENSITY

Strong gravitational lensing occurs when the surface mass density along a given sightline exceeds a certain critical value, and is associated with high magnifications, multiple images, Einstein arcs and rings in the lens plane.

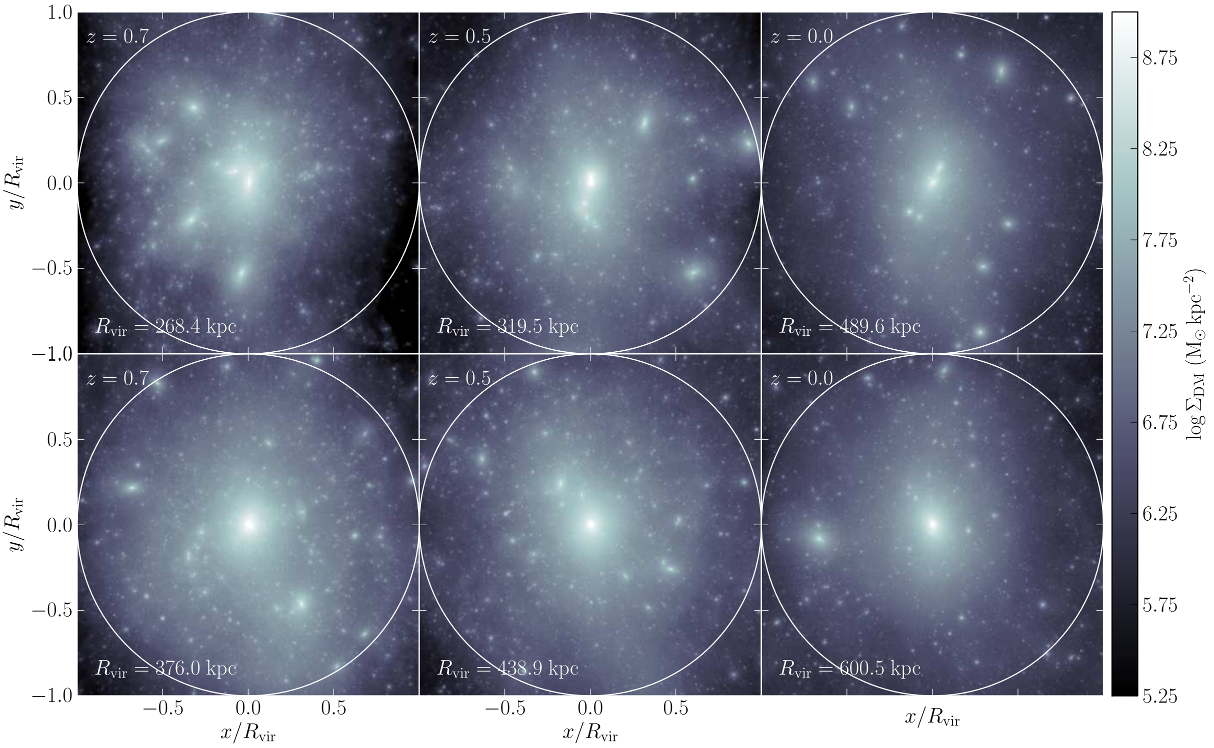

Figure 6. Projected dark matter density at $z = 0, 0.5, 0.7$ (from right to left) of PonosQ (top) and PonosV (bottom) in a slice of thickness $2R_{\rm vir}$ centered on the target halos. The image brightness is proportional to the logarithm of the total dark matter surface density along the line-of-sight, $\Sigma_{\rm DM}$. The circles mark the virial radius $R_{\rm vir}$.

Figure 6 shows the surface dark matter mass density at $z = 0, 0.5, 0.7$ of PonosQ and PonosV in a slice of thickness equal to $2R_{\rm vir}$ centered on the target halos. As expected in ΛCDM, the two Ponos halos are teeming with self-bound subhalos on all resolved mass scales. To shed light on whether CDM substructures can account for the observed lensing flux ratio anomalies, we have measured in the simulations the surface mass fraction in subhalos within 20 different azimuthal annuli equally spaced in $\log(R/R_{\rm vir})$ from −2 to 0,

$$F_{\rm sub}(R) = \frac{\Sigma_{\rm sub}(R)}{\Sigma_{\rm DM}(R)}, \quad (5)$$

where R is the projected radius measured from the host center, and $\Sigma_{\rm sub}$ and $\Sigma_{\rm DM}$ are the substructure and total dark matter projected mass densities, respectively. The surface density in substructures is computed by summing up all the subhalo bound mass that falls within the relevant annulus. To be specific, we adopt the following procedure to compute $F_{\rm sub}$. First, we determine a line-of-sight by sampling uniformly the solid angle 4π. This is achieved by randomly drawing the azimuthal angle ϕ from a uniform distribution in the interval $[0, 2\pi)$, and the co-latitudinal angle θ from the distribution $\mathcal{P}(\theta) = \sin(\theta)/2$. Then, we determine the line-of-sight through the versor $\mathbf{n} = (\sin\theta\ \cos\phi, \sin\theta\ \sin\phi, \cos\theta)$. We calculate the projected distance of a particle in the plane perpendicular to $\mathbf{n}$ as $R = \sqrt{|\mathbf{r}|^2 - (\mathbf{r}\cdot\mathbf{n})^2}$, where $\mathbf{r}$ is the position vector of the particle from the centre of the halo, and we finally bin the positions of each particle, adding its mass to the proper radial bin. We repeat this for both the subhalos only and the host halo, and we finally calculate the ratio between the two to derive $F_{\rm sub}(R)$.

Figure 7 shows the *median* subhalo surface mass fraction at $z = 0, 0.5, 0.7$ as a function of $R/R_{\rm vir}$ over 300 random projections for each of the two Ponos halos. The 68% scatter among the different projections is marked by the colored bands, and is larger at smaller radii as the area of the azimuthal annulus around R becomes progressively smaller in the inner regions. The value of $F_{\rm sub}$ is clearly sensitive to the assembly history of the host galaxy. The black dashed line depicts the results derived by Mao et al. (2004) at $z = 0$ from simulations of twelve halos of galactic, group, and cluster masses. Their surface mass fraction in substructures is somewhat lower than found here, perhaps because of resolution limitations – the number of particles per halo in their numerical investigations is about two orders of magnitudes smaller than here.

The statistical study by Dalal & Kochanek (2002) of the anomalous flux ratios observed in a sample of seven lensed radio-loud quasars requires $F_{\rm sub}$ =0.6 to 7% (90% confidence, with a median of 2%) of the mass at the Einstein radius to be in substructures in order to reproduce the data. The black point in the figure indicates the Dalal & Kochanek (2002) constraint – for an assumed

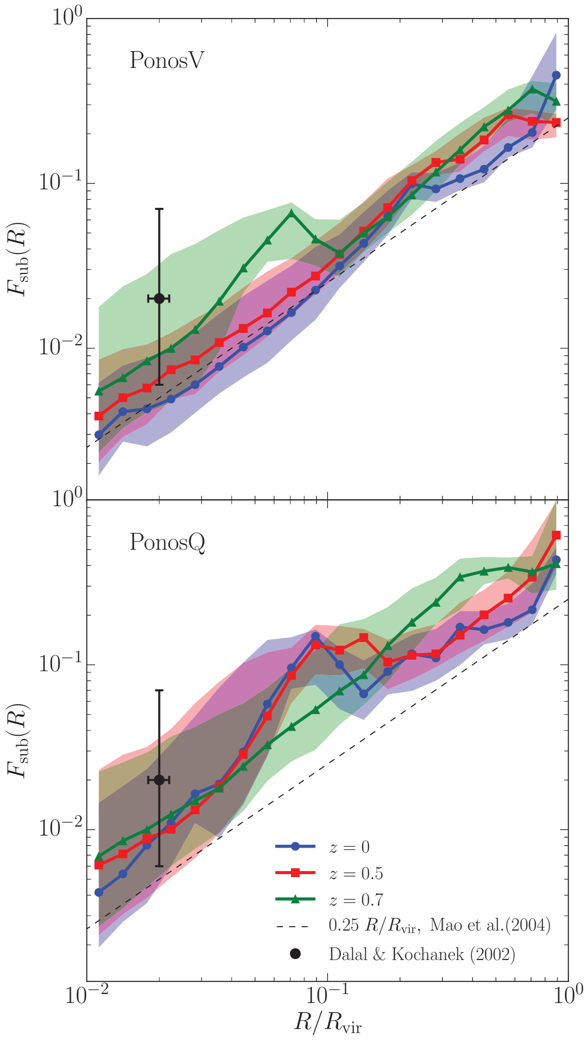

Figure 7. Predicted substructure surface mass fraction at $z = 0, 0.5, 0.7$ in PonosV (top panel) and PonosQ (bottom panel). The panels show $F_{\rm sub}$ in azimuthal annuli as a function of the projected radius, R (in units of the virial radius), for all subhalos belonging to the corresponding host. The solid lines depict the median over 300 random projections for each of the two Ponos halos, while the colored bands mark the 68% scatter among the different projections. The black point indicates the median and 90% confidence level of the mass fraction in local substructure required to explain the flux ratio anomalies (for an assumed Einstein radius of $0.02\,R_{\rm vir} \sim$5–10 kpc). The dashed black curve shows the results derived by Mao et al. (2004) at $z = 0$ from low resolution simulations of twelve halos of galactic, group, and cluster masses.

Einstein radius of $0.02\,R_{\rm vir} \sim$ 5–10 kpc, which appears entirely consistent with the expectations for ΛCDM. At the Einstein radius, the *mean* surface mass fraction in substructure for the two Ponos halos ranges from 0.6% to 2.3%. In contrast with PonosV, there is hardly any redshift dependence in the value of this quantity for PonosQ.

These projected mass fractions are considerably higher than the mean value of 0.2% ($0.01 \le F_{\rm sub} \le 0.7\%$) measured at the Einstein radius in the Aquarius simulations of Milky Way-sized halos (Xu et al. 2009). Recognizing that the substructure abundance in group-sized halo may be larger than in less massive Milky Way-sized hosts, Xu et al. (2015) recently rescaled the subhalo populations of the Aquarius suite of galactic halos and the Phoenix suite of cluster halos to those expected in massive ellipticals. They estimate a mean surface mass fraction of substructure at the Einstein radius that is three times bigger than in Milky Way-sized hosts, in better agreement with our findings. We note, however, that PonosQ has a mean subhalo surface mass fraction that is still a factor of ~ 2 above the determinations by Xu et al. (2015).

At $z = 0$, we find a mean substructure surface mass density around the Einstein radius of $\Sigma_{\rm sub} = 3.6 \times 10^6\,{\rm M}_\odot\,{\rm kpc}^{-2}$ in PonosV, and $8.9 \times 10^6\,{\rm M}_\odot\,{\rm kpc}^{-2}$ in PonosQ. At $z = 0.7$, we measure $\Sigma_{\rm sub} = 1.4 \times 10^7\,{\rm M}_\odot\,{\rm kpc}^{-2}$ and $1.0 \times 10^7\,{\rm M}_\odot\,{\rm kpc}^{-2}$ in PonosV and PonosQ, respectively. Let us denote with η the mean column number density of subhalos at projected halocentric distance $R = 0.02\,R_{\rm vir}$. Over the interval $10^8\,{\rm M}_\odot < M_{\rm sub} < 10^{11}\,{\rm M}_\odot$, the mean projected mass function can be written as

$$\frac{d\eta}{d\ln M_{\rm sub}} = \mathcal{N}_0 \left(\frac{M_{\rm sub}}{m_0}\right)^{-\ell}. \qquad (6)$$

At $z = 0.7$, we measure in PonosV a best-fit slope of $\ell = 0.85$ and a column number density at the pivot mass, $m_0 = 10^8\,{\rm M}_\odot$, of $\mathcal{N}_0 = 0.006\ {\rm kpc}^{-2}$. In PonosQ we find a similar normalization but a steeper slope, $\ell = 1.09$. At $z = 0$, we measure $\ell = 0.98$ in PonosV and $\ell = 0.84$ in PonosQ. Most subhalos at these image positions only appear along the line-of-sight because of projection effects.

In Figure 8 we plot the *mean* and *median* surface mass fraction per decade of subhalo mass, $\Delta F_{\rm sub}$, again at projected halocentric distance $R = 0.02\,R_{\rm vir}$. As expected from a subhalo projected mass function $d\eta/d\ln M_{\rm sub} \propto M_{\rm sub}^{-\ell}$ with ℓ close to 1, this distribution is relatively flat, with roughly equal contributions per decade of mass above the "completeness" mass scale of $M_{\rm sub} = 10^7\,{\rm M}_\odot$. The contribution of subhalos with $M_{\rm sub} < 10^9\,{\rm M}_\odot$ to the projected substructure mass fraction is between one fifth and one third of the total, with the smallest share found in the quiescent host. Massive subhalos with $M_{\rm sub} \gtrsim 10^{10}\,{\rm M}_\odot$ only survive in the outer, $\langle r\rangle \gtrsim 160\,{\rm kpc}$, regions of their hosts because of tidal destruction. Their presence in the projected central $\sim$ 10 kpc is then typically associated with chance alignment. This explains the large disparity between the mean and median $\Delta F_{\rm sub}$ in the largest mass bins. The insets in the figure show the broad probability density function (PDF) of the total $F_{\rm sub}$ around $0.02R_{\rm vir}$, for the 300 lines of sight. At $z = 0.7$, the median $F_{\rm sub}$ is about 1%, and the probability of observing values of $F_{\rm sub} > 0.05$ can be as large as 12% (PonosQ).

5. BARYONIC CONTRACTION

The most severe limitation of our study is that the N-body very high resolution simulations used here include only dark matter. Standard dark matter halos are poor lenses because their central cusps ($\rho \propto r^{-1}$) are too shallow. In the adopted cosmology, and for a typical lens geometry with $z_{\rm lens} = 0.7$ and $z_{\rm source} = 2$, the critical surface density for multiple imaging is $\Sigma_{\rm crit} = 2.2 \times 10^9\,{\rm M}_\odot\,{\rm kpc}^{-2}$. At the same redshift and $R = 0.02\,R_{\rm vir}$, our Ponos lens halos have $\Sigma/\Sigma_{\rm cr} = 0.2 - 0.4$, i.e. are subcritical and unable to produce multiple images despite

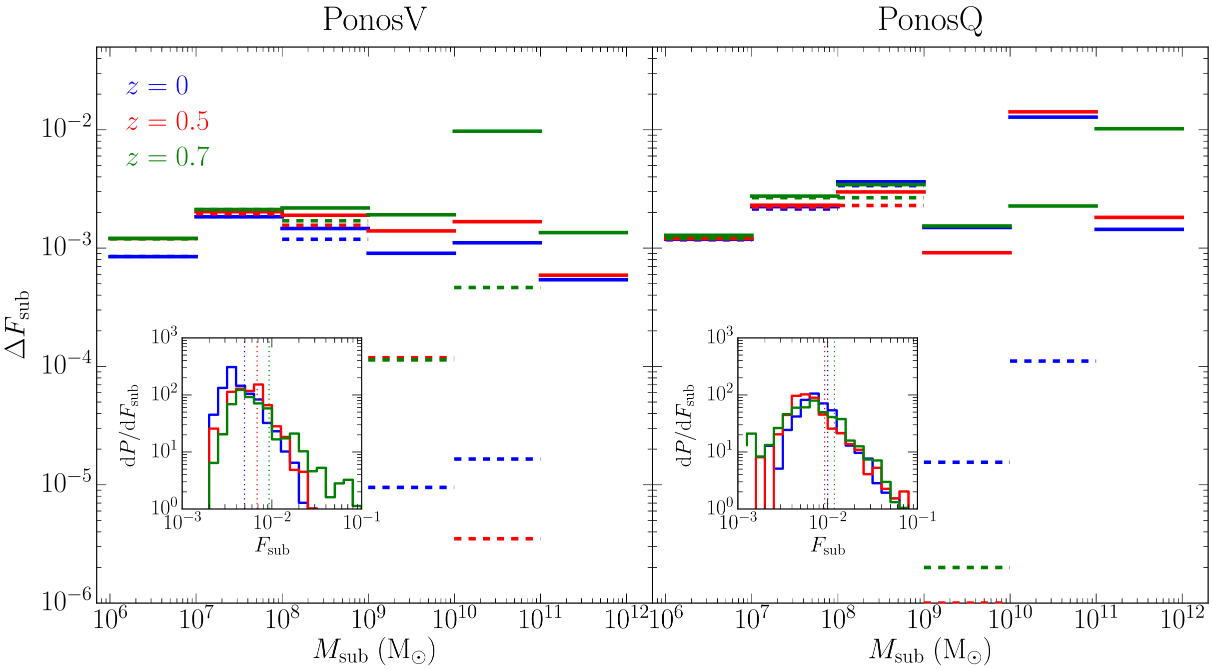

Figure 8. Projected substructure mass fraction at $R = 0.02\,R_{\rm vir}$ in each mass decade, for PonosV (left panel) and PonosQ (right panel). The individual horizontal bars depict the mean (solid lines) and median (dashed lines) over 300 random projections for each host. The three different redshifts inspected are listed in the left) panel. The insets show the broad probability density function of the total $F_{\rm sub}$, for the 300 lines of sight, at the three redshifts. Median values are indicated by the vertical dotted lines.

having maximum circular velocities close to $240\,{\rm km\,s^{-1}}$ (PonosQ) and $340\,{\rm km\,s^{-1}}$ (PonosV). Baryon cooling and condensation is expected to strongly affect the inner density profiles of halos, although predictions for these effects are far less certain than forecasting the distribution of dark matter in the purely dissipationless regime. As the baryons cool, they drag some of the dark matter inward and may even dominate the mass within one Einstein radius, thereby converting a sub-critical dark matter halo into one capable of producing multiple images.

Rather than putting in baryons "by hand", for example by implanting an idealized model galaxy into a dark matter halo extracted from a collisionless simulation (Amara et al. 2006) or by modeling the main lens halo as a singular isothermal ellipsoid (e.g., Metcalf & Amara 2012; Xu et al. 2015), we have followed here a different approach and run a lower resolution version of PonosQ ("PonosQH") with hydrodynamics using the TreeSPH code GASOLINE (Wadsley et al. 2004). The code employs a subgrid model for the turbulent mixing of metals and energy following Shen et al. (2010). The ICs for PonosQH were initialized as for the collisionless run. The high-resolution region contains 6 million dark matter particles and an equal number of gas particles, for a dark matter and (initial) gas particle mass of $m_{\rm DM} = 2.3 \times 10^6\,{\rm M_\odot}$ and $m_{\rm SPH} = 4.5 \times 10^5\,{\rm M_\odot}$, respectively. This mass resolution is very similar to that of the "small group mass" halo in the FIRE simulation suite (Hopkins et al. 2014). The gravitational softening length was fixed to 785 pc (physical) for the dark matter, and to 501 pc for the gas. In high-density regions the gas smoothing length was allowed to shrink to 10% of the softening to ensure that hydrodynamic forces are resolved on 100 pc scales. A non-thermal pressure floor (Agertz et al. 2009; Roškar et al. 2015) was applied to stabilize scales of order the gravitational softening against gravitational collapse and avoid artificial fragmentation.

The simulation includes a non-equilibrium primordial chemistry network for atomic H and He, Compton cooling off the cosmic microwave background, and precomputed tabulated metal-line cooling rates from the photoionization code CLOUDY (Ferland et al. 1998). A spatially-uniform, redshift-dependent cosmic UV background (Haardt & Madau 2012) modifies the ionization and excitation state of the gas, photoionizing away abundant metal ions and reducing the cooling efficiency. Star formation proceeds at a rate $d\rho_\star/dt = 0.05(\rho_{\rm gas}/t_{\rm dyn}) \propto \rho_{\rm gas}^{3/2}$ (i.e., locally enforcing the Schmidt law), where $\rho_\star$ and $\rho_{\rm gas}$ are the stellar and gas densities, and $t_{\rm dyn}$ is the local dynamical time. Star particles form in cold gas (i.e. temperature below 10^4 K) that reaches a density threshold of 20 atoms ${\rm cm^{-3}}$, and are created stochastically with an initial mass $m_\star = 1.35 \times 10^5\,{\rm M_\odot}$ distributed following a Kroupa (2001) initial mass function. They inject energy, mass, and metals back into the interstellar medium (ISM) through Type Ia and Type II supernovae (SNe) and stellar winds, following the prescriptions of Stinson et al. (2006). A "delayed radiative cooling" scheme for Type II SN feedback was adopted, a simplified algorithm designed to extend the Sedov-Taylor phase of SN remnants and mimic the effect of energy deposited in the local ISM by multiple, clustered sources of mechanical luminosity (Stinson et al. 2006). While this approach has been found to be key in reproducing the properties of dwarfs (Governato et al. 2010; Madau et al. 2014; Shen et al. 2014), late-type spirals (Guedes et al. 2011), and the circumgalactic medium of $z \sim 3$ galaxies (Shen et al.

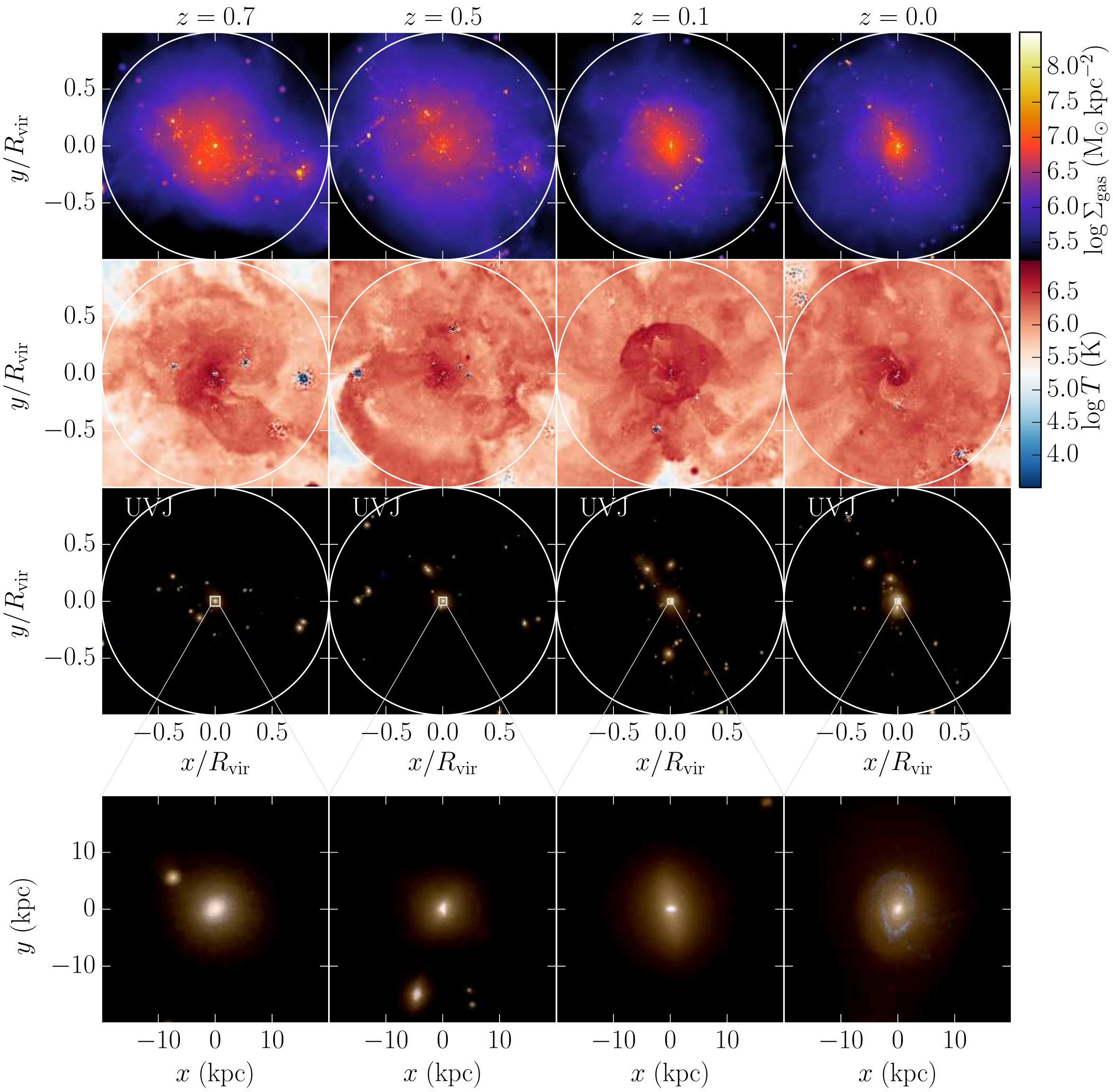

Figure 9. The gas and stellar properties of PonosQH at different redshifts. From top to bottom: gas surface-density, gas temperature in a slice through the center of the host, mock images in the U, V, and J filter bands. The top three panels encompass a physical scale equal to the virial radius, while the bottom panels shows images of the inner 40 kpc (physical).

2013), many authors have recently stressed the importance of correctly accounting for the entire momentum budget of stellar feedback, including momentum injection by radiation pressure, stellar winds, and clustered SN explosions (see, e.g. Agertz et al. 2013; Hopkins et al. 2014; Keller et al. 2014; Kim & Ostriker 2015). It has been shown by Agertz et al. (2013) that simulations with maximal momentum injection suppress star formation to a similar degree than found in runs that, like ours, adopt adiabatic thermal feedback. We note, however, that our stellar feedback model is not strictly speaking adiabatic (since SN-heated gas particles exchange thermal energy with the ambient medium via turbulent mixing and therefore can cool "indirectly"), and that we did not include any feedback from a central active galactic nucleus (AGN).

Figure 9 shows gas surface-density and temperature maps at four different redshifts, together with mock images in the U, V, and J filter bands. The top three panels encompass a physical scale equal to the virial radius, while the bottom panels are images of the inner 40 kpc (physical). Each star particle was assigned a luminosity from tables of mass-to-light ratios based on the isochrones and synthetic stellar populations of Bressan et al. (2012). At the present epoch, PonosQH has an early-type morphology, with $U - V = 1.5$ and $V - J = 0.9$ colors that are typical of red sequence galaxies. Its present-day stellar mass within 20 kpc from the center, $M_\star = 2.3 \times 10^{11}\,\rm M_\odot$, implies a star formation efficiency, $M_\star/M_{\rm vir} = 0.035$, which is slightly higher than the value inferred for massive halos by Kravtsov et al. (2014), but compatible within 2σ. The surface brightness profile fol-

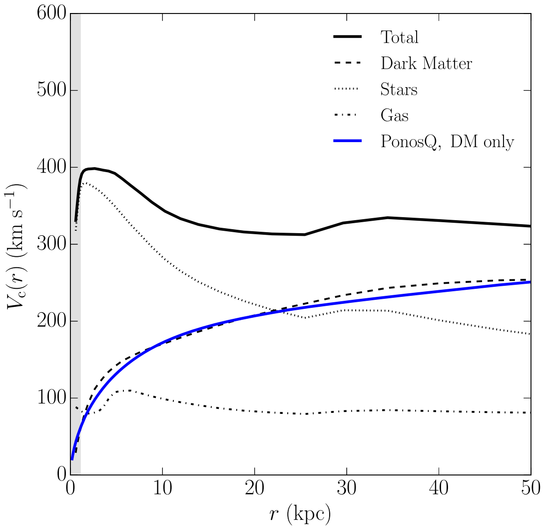

Figure 10. The $z = 0$ rotation curve of PonosQH, for all components (black solid line), stars (dotted line), gas (dash-dotted line), and dark matter (dashed line). For comparison, the dark matter rotation curve (blue solid line) obtained in the purely collisionless simulation of the same object is also shown. The gray shading indicates the region inside two gravitational softening lengths of the baryonic component. The bump at $\sim$ 35 kpc from the center is caused by a massive satellite system.

lows a de Vaucouleurs law with a V-band effective radius $R_{e,V} = 3.6$ kpc, while the half-light radius measured directly from the light profile is $R_{1/2} = 3.1$ kpc. The baryonic content of PonosQH is close to the universal baryon fraction. The mass of cold ($T < 10^4$ K) gas per unit K-band luminosity, $M(\mathrm{HI})/L_K = 0.03$, places this system towards the high-end of the distribution observed in the ATLAS survey of nearby early-type galaxies (Serra et al. 2012).

We now focus on the the impact of baryonic infall on the inner density profile and projected substructure mass fraction. In Figure 10 we show the $z = 0$ rotation curve, $V_c(r) = \sqrt{GM(<r)/r}$, for all components (gas, stars, and dark matter) separately. The dark matter circular velocity curve obtained in the collisionless simulation of the same object is plotted for comparison. Baryons dominate within the inner 20 kpc, and the ensuing rotation curve remains relatively flat from several kpc to several tens of kpc. The bump at $\sim$ 35 kpc is due to the presence of a massive, $\sim 1.5 \times 10^{11} \mathrm{M}_\odot$, satellite. The total mass within 10 kpc has increased by a factor of three relative to the purely collisionless simulation. The effective power-law slope of the total (luminous plus dark; $\rho \propto r^{-\gamma}$) mass distribution in the range 1.5-40 kpc is $\gamma = 2.19 \pm 0.07$, i.e. marginally steeper than isothermal ($\gamma = 2$). This value appears to be consistent with the results of the Sloan Lens ACS Survey (SLACS) of 73 early-type galaxies with $0.08 < z_{\mathrm{lens}} < 0.5$ and stellar masses above $10^{11}\,\mathrm{M}_\odot$: $\langle\gamma\rangle = 2.078 \pm 0.027$ with an intrinsic scatter of 0.16 ± 0.02 (Auger et al. 2010). Recent modeling of the mass density profiles of early-type galaxies also yields super-isothermal central slopes, with $\langle\gamma\rangle = 2.15 \pm 0.04$ (Chae et al. 2014) and $\langle\gamma\rangle = 2.19 \pm 0.03$ (Cappellari et al. 2015).

The steeper inner cusp is clearly seen in the density profile plot (Figure 11). In response to the slow addition of baryons to the center, dark matter may be pulled inwards through a process known as adiabatic contraction (Blumenthal et al. 1986; Gnedin et al. 2004; Pillepich et al. 2014). Other processes may cause the dark matter to expand instead, such as the transfer of orbital energy via dynamical friction following dry minor mergers (El-Zant et al. 2001) and rapid gravitational poten-

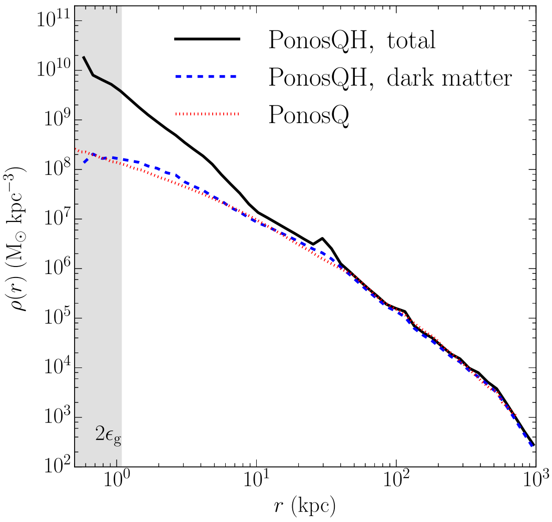

Figure 11. The $z = 0$ density profile in PonosQH for all components (black solid line) and dark matter (dashed line). For comparison, the dark matter density profile (red dotted line) obtained in the purely collisionless simulation of the same object is also shown. The gray shading indicates the region inside two gravitational softening lengths of the baryonic component. We do not observe any significant evidence for adiabatic contraction or expansion in the dark matter at the center of the halo in the hydrodynamical run.

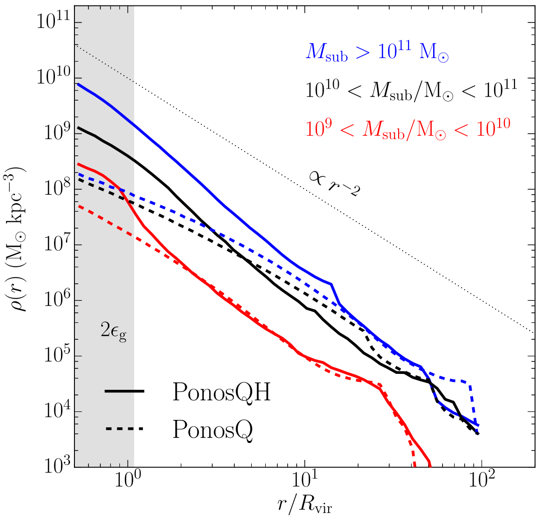

Figure 12. The $z = 0$ average density profiles of massive subhalos. Solid lines: PonosQH. Dashed lines: PonosQ. The difference between the two sets shows the impact of baryonic contraction. The red, black, and blue colors refer to the subhalo mass intervals $10^9 < M_{\mathrm{sub}}/\mathrm{M}_\odot < 10^{10}$, $10^{10} < M_{\mathrm{sub}}/\mathrm{M}_\odot < 10^{11}$, and $M_{\mathrm{sub}}/\mathrm{M}_\odot > 10^{11}$, respectively. For reference, the dotted line shows an isothermal, $\propto r^{-2}$, profile. The gray shading indicates the region inside two gravitational softening lengths of the baryonic component.

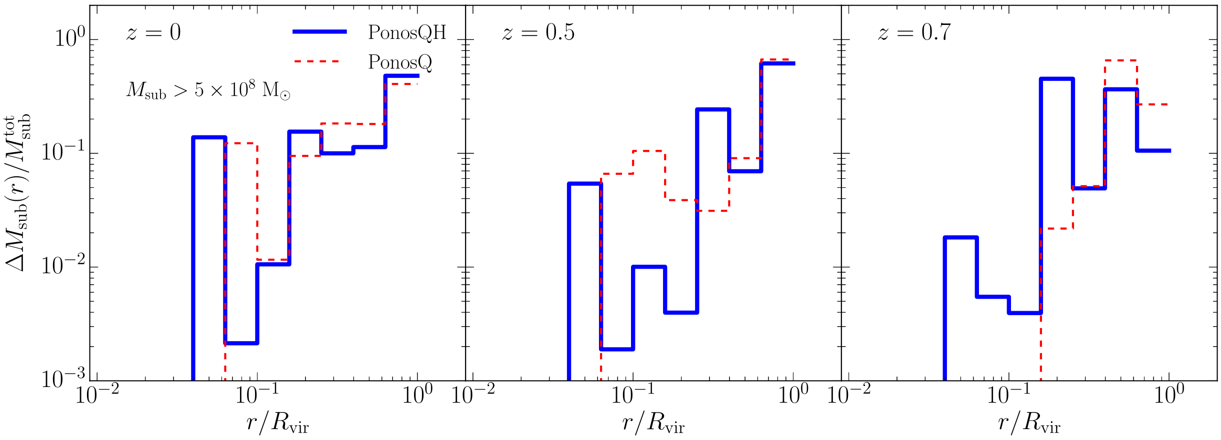

Figure 13. Impact of baryonic contraction on the radial distribution of massive subhalos. Solid line: radial distribution of the cumulative mass, $\Delta M_{\rm sub}$, in all PonosQH subhalos with $M_{\rm sub} > 5 \times 10^8\,{\rm M}_\odot$. Dashed line: same for PonosQ. The distance to the center of the main halo is normalized to the virial radius. The subhalo mass is normalized to the total mass in subhalos $> 5 \times 10^8\,{\rm M}_\odot$ identified in the hydro and collisionless simulations. The left, middle, and right panels show the distribution at three different redshifts.

tial fluctuations tied to efficient SN feedback (Pontzen & Governato 2012; Madau et al. 2014). At the present-epoch, PonosQH shows little evidence for strong dark matter contraction or expansion. Using a generalized NFW model, we find an inner dark matter density slope, $\gamma_{\rm DM} = 1.1 \pm 0.1$, which is consistent with unity. The mean dark matter mass fraction projected within a cylinder of radius equal to the Einstein radius ($0.02\,R_{\rm vir}$) is $F_{\rm DM}(< R_{\rm vir}) = 0.38 \pm 0.01$, in agreement with the average value, 0.38 ± 0.07, measured in the SLACS lens sample by Bolton et al. (2008).

Baryonic contraction acts both for the host halo and for its subhalos. At $z = 0$, we identify in PonosQH more than 400 subhalos, 285 of which are above $10^8\,{\rm M}_\odot$ (the minimum subhalo mass corresponding to 20 bound dark matter particles is now $M_{\rm res} = 4.6 \times 10^7\,{\rm M}_\odot$). Figure 12 shows the $z = 0$ density profiles of massive, $M_{\rm sub} > 10^9\,{\rm M}_\odot$ subhalos averaged over different mass intervals. All subhalos above $M_{\rm sub} > 10^{10}\,{\rm M}_\odot$ form stars and are therefore "luminous", while half of those in the range $10^9 < M_{\rm sub}/{\rm M}_\odot < 10^{10}$ are dark. The difference between PonosQH and the collisionless PonosQ is clear. Baryonic infall increases the total mass within, say, 2 kpc from the satellite center by a factor that ranges from 2 ($10^9 < M_{\rm sub}/{\rm M}_\odot < 10^{10}$) to 14 ($M_{\rm sub}/{\rm M}_\odot > 10^{11}$) relative to the purely collisionless simulation. As a consequence, the mean maximum circular velocity of subhalos in the mass range $10^9 < M_{\rm sub}/{\rm M}_\odot < 10^{10}$ rises from $28\,{\rm km\,s^{-1}}$ to $49\,{\rm km\,s^{-1}}$, and from $83\,{\rm km\,s^{-1}}$ to $122\,{\rm km\,s^{-1}}$ in the case of $10^{10} < M_{\rm sub}/{\rm M}_\odot < 10^{11}$ subhalos. Similar effects of baryonic contraction in massive satellites of Milky Way-sized halos have also been recently reported by Zhu et al. (2015). And while the slope of the internal density profile of subhalos may have little effect on the frequency of flux anomalies (Metcalf & Amara 2012), baryon cooling and condensation within massive subhalos will make them more resilient to tidal disruptions (e.g. Macciò et al. 2006).

In Figure 13 we compare the radial distribution of the cumulative subhalo mass in PonosQH and PonosQ. Only subhalos with $M_{\rm sub} > 5 \times 10^8\,{\rm M}_\odot$ are included in the comparison. The distance to the center of the main halo is normalized to the virial radius (slightly larger in PonosQH), while the subhalo mass is normalized to the total mass in all $M_{\rm sub} > 5 \times 10^8\,{\rm M}_\odot$ subhalos identified in the hydro and collisionless simulations. The left, middle, and right panels show subhalos at three different redshifts. The substructure mass in the inner regions is consistently larger in PonosQH. This notwithstanding the fact that the increased mass concentration at the center of the main host induces stronger tidal forces that can potentially destroy subhalos. At $z = 0.7$, for example, there are 4 subhalos more massive than $10^9\,{\rm M}_\odot$ within 30 kpc of the center of PonosQH, compared to none in PonosQ. At $z = 0.5$, the innermost $M_{\rm sub} > 10^{10}\,{\rm M}_\odot$ satellite in PonosQH is at 17 kpc from the center, versus 46 kpc for PonosQ.

The effect of baryon cooling and condensation on the predicted projected mass fraction is depicted in Figure 14. We have modified the numerator and denominator of Equation (5) to account for the impact of baryons on the main halo and subhalos as

$$F_{\rm sub}(R) = \frac{\Sigma_{\rm sub}(R)}{\Sigma_{\rm DM}(R) + \Sigma_{\rm bar}(R)}. \tag{7}$$

Here, $\Sigma_{\rm sub}$ is the substructure projected mass (baryons + dark matter) density, while $\Sigma_{\rm DM}$ and $\Sigma_{\rm bar}$ are the total dark matter and baryon surface densities of the main host, all measured in PonosQH. *PonosQH does not have enough resolution to generate a realistic substructure population at masses $M_{\rm sub} \lesssim 5 \times 10^8\,{\rm M}_\odot$, and so this figure should be interpreted as the surface mass fraction in substructure at intermediate and large mass scales only.* As in Figure 7, we have plotted the median substructure mass fraction rather than the mean, for better comparison with the Dalal & Kochanek (2002) constraints. Note how, even without small-scale power, the predicted median at the Einstein radius and $z \gtrsim 0.5$ is consistent with the data.

6. DISCUSSION AND CONCLUSIONS

In this Paper, we have used new very high resolution N-body and hydrodynamical cosmological simulations of

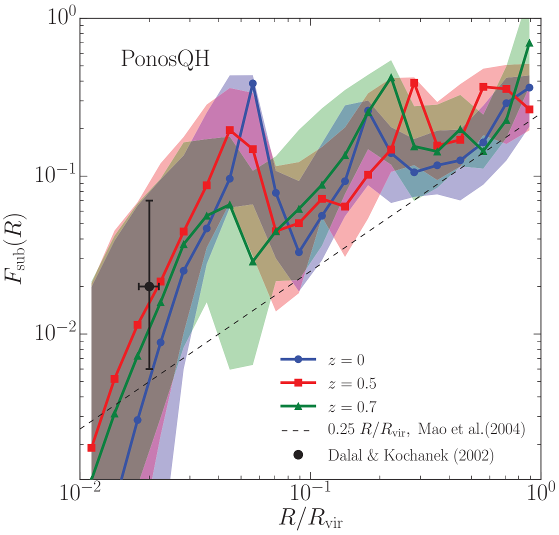

Figure 14. Impact of baryonic contraction on the substructure surface mass fraction. Here, we have measured the total (baryons + dark matter) substructure surface density in the hydro simulation PonosQH. Legend is the same as Figure 7. Because PonosQH only resolve substructures at intermediate and large mass scales, the sightline-to-sightline scatter has increased compared to Figure 7.

early-type massive galaxies to refine ΛCDM predictions for substructure gravitational lensing. It has long been known that, if dark matter substructure can survive to constitute a percent or more of the surface mass density at small impact parameters, it will cause detectable anomalies in the magnification ratios of multiply-imaged QSOs (Metcalf & Madau 2001). In our dissipationless simulations, we have found that the total mass fraction in self-bound subhalos increases from about 15% at the present epoch to 20-30% at redshift 0.7. In projection, the average fraction of surface mass density in substructure around the Einstein radius at $z = 0.7$ exceeds 2%. More massive hosts (as well as hosts at higher redshifts) are predicted to be dynamically younger and therefore more clumpy, as their subhalos are accreted more recently and tend to survive tidal destruction. Indeed, our two Ponos $M_{\rm vir} \sim 10^{13}\,{\rm M}_\odot$ halos have significantly higher, by as much as a factor of 10, mean projected substructure mass fractions at $z = 0.7$ than measured at the present-epoch in the Aquarius and Via Lactea simulations of Milky Way-sized systems. As a result, the frequency of flux ratio anomalies predicted by the richer substructure population of early-type galaxy halos increases noticeably compared to estimates based on Milky Way-sized halos (e.g., Amara et al. 2006; Macciò et al. 2006; Xu et al. 2009).

We have incorporated the effects of baryonic contraction on the host halo using twin, lower-resolution hydrodynamical simulations that include metallicity-dependent gas cooling, a star formation recipe based on a high gas density threshold, and a delayed-radiative-cooling scheme for feedback by SNe, but no AGN feedback. The inclusion of the baryonic component produces at the present epoch a red sequence galaxy with a super-isothermal central slope in the total (luminous plus dark) mass distribution, a rotation curve that is relatively flat from several kpc to several tens of kpc, a projected dark matter mass fraction inside the Einstein radius of 40%, and little evidence for strong dark matter contraction or expansion. Such properties of the matter density profile appear broadly consistent with observations of early-type galaxies (e.g., Bolton et al. 2008; Auger et al. 2010; Chae et al. 2014; Cappellari et al. 2015). Baryonic contraction increases the number of massive subhalos in the inner regions of the main host.

Our ΛCDM simulations, both in the purely collisionless case or after accounting for the impact of baryonic contraction, appear at first sight to predict enough substructure to explain the frequency of lens anomalies in currently available samples (Dalal & Kochanek 2002). While this is indeed promising, numerical calculations of the lensing potential, deflection angles, and magnifications are needed to generate theoretical flux-ratio probability distributions for comparison with the observations. Two such studies have been performed recently. Metcalf & Amara (2012) created a large number of simulated lenses with finite source sizes, compared the predicted flux-ratio probability distributions in the presence of substructure to an observational sample of seven lenses, and found approximate consistency with ΛCDM N-body simulations. Xu et al. (2015) constructed samples of lens potentials by adding (rescaled to those expected in massive ellipticals) subhalo populations from the galaxy-scale Aquarius and the cluster-scale Phoenix simulation suites, and matched the resulting flux ratio distributions to the best available sample of radio lenses. They reached the conclusion that CDM substructures cannot account for all the observed anomalies. Such detailed investigations are beyond the scope of this work, and we defer them to a future paper. Here, we only note that, according to Figure 9 in Metcalf & Amara (2012), a subhalo surface mass density of $\Sigma_{\rm sub} = 10^7\,{\rm M}_\odot\,{\rm kpc}^{-2}$ would imply a 15% chance of a clear outlier in the general distribution of $\Delta\theta$ and $R_{\rm cusp}$ values predicted in the absence of substructures, consistent with the frequency (one out of seven) actually observed[9]. A substructure surface mass density of $10^7\,{\rm M}_\odot\,{\rm kpc}^{-2}$ is comparable to that measured in the Ponos halos. Contrary to the model of Metcalf & Amara (2012), however, most of this projected mass density is associated with subhalos more massive than $M_{\rm sub} = 10^9\,{\rm M}_\odot$.

Which brings up the next point. Magnifications, which depend on the second spatial derivative of the lensing potential, are affected about equally by all mass scales provided the Einstein radius of the deflector is larger than the size of the background source. The constraints provided by the frequency of flux anomalies cannot therefore discriminate between different substructure mass scales. On the other hand, distinguishing CDM from keV-mass WDM – in which the free-streaming cutoff occurs on dwarf galaxy scales – requires measuring the subhalo

---

[9] Here, $\Delta\theta$ and $R_{\rm cusp}$ are two parameters that characterize 4-image lenses. Sources near a cusp in the caustic produce "cusp" configurations with three of the images ("triplet") lying close together on one side of the lens galaxy. The parameter $\Delta\theta$ is the angular separation between the close triplet, and is small when the source is near one of the cusps in the caustic. In any smooth lensing potential, the three close images satisfy an asymptotic magnification relation (the "cusp-caustic relation") $R_{\rm cusp} = |\mu_A + \mu_B + \mu_C|/(|\mu_A| + |\mu_B| + |\mu_C|) \to 0$, with the total absolute magnification $|\mu_A| + |\mu_B| + |\mu_C| \to \infty$ (e.g. Keeton et al. 2003).

mass function well below a mass of $\sim 10^9\,\mathrm{M}_\odot$ (e.g., Li et al. 2015). We find that, in the Ponos hosts, the contribution of subhalos with $M_{\rm sub} < 10^9\,\mathrm{M}_\odot$ to the total projected mass fraction is sub-dominant, between one fifth and one third of the total. This fact highlights the potential importance of dwarf-sized systems in the anomalous flux ratio problem. Small dark substructures may actually play a lesser role in causing flux anomalies than the more massive subhalos, and some of these massive perturbers may actually contain visible dwarf galaxies. Indeed, three out of the seven gravitational lens systems used by Dalal & Kochanek (2002) to statistically detect dark substructure around early-type galaxies show evidence for additional mass structure in the form of luminous dwarf satellites (see, e.g., McKean et al. 2007, and references therein). In the case of the lens systems B2045+265 and MG 2016+112, the dwarf perturber constitutes about 1% of the total projected lens mass, which seems consistent with the predictions of our Ponos simulations. Moreover, other components of the lensing galaxy, like an edge-on disk, may also account for some anomalies, as suggested by recent observations of CLASS B1555+375 (Hsueh et al. 2016). It remains therefore unclear how effective a probe of the matter power spectrum on sub-galactic scales the current, small sample of lensed radio-loud quasars may actually be. Another complicating factor makes the comparison between ΛCDM predictions with the observations still rather preliminary. In addition to substructures within the halo of the lensing galaxy, dwarf-sized perturbers along the line-of-sight to the lensed quasar may also affect the lens potential and give rise to flux ratio anomalies. And while lensing by dark matter clumps near the lens galaxy may be more effective, the cumulative effect of all intergalactic halos along the line-of-sight could be significant (Chen et al. 2003; Wambsganss et al. 2005; Metcalf 2005; Inoue & Takahashi 2012; Inoue 2016).

With the next generation of wide-field optical surveys capable, e.g., of increasing the samples of multiply-imaged quasars by two orders of magnitude (Oguri & Marshall 2010), however, strong gravitational lensing appears poised to dramatically improve our understanding of dark matter substructure in galaxy halos. Flux ratio anomalies are only one means of detecting substructure. The most massive subhalos may also visibly perturb the deflection angle of lensed images, causing "astrometric anomalies" (Metcalf & Madau 2001; Chiba 2002; Chen et al. 2007). High-precision measurements of time delay perturbations between images in strong gravitational lens systems may complement flux ratio and astrometric anomalies as they depend on a different moment ($M_{\rm sub}^2$) of the subhalo mass function (Keeton & Moustakas 2009). An alternative technique, the "direct gravitational imaging" of individual clumps based on perturbations to the surface brightness of highly magnified Einstein rings, has been developed by Koopmans (2005). Detailed studies of individual Einstein ring systems have led Vegetti et al. (2012) to detect a $(1.9 \pm 0.1) \times 10^8\,\mathrm{M}_\odot$ dark subhalo in an extended optical galaxy-galaxy lens system at $z = 0.88$. From a search of mass clumps in a sample of 11 lens galaxies from the SLACS, Vegetti et al. (2014) infer a mean projected substructure mass fraction $F_{\rm sub} = 0.0076^{+0.0208}_{-0.0052}$ (68% confidence level) around the Einstein radius of massive early-type host galaxies at $\langle z_{\rm lens}\rangle = 0.2$. This fraction is again consistent with the expectations from ΛCDM simulations. Recently, Hezaveh et al. (2016, see also Inoue et al. 2016) have used ALMA observations of the strongly lensed dusty galaxy SDP.81 to find evidence for a $M_{\rm sub} = 10^{8.96\pm0.12}\,\mathrm{M}_\odot$ subhalo near one of the images, and produce constraints on the projected abundance of substructure near the Einstein radius. Figure 15 shows the resulting constraints on the differential projected subhalo mass function, compared with predictions from the Ponos simulation sets. Again, the observations appear consistent with theoretical expectations. More importantly, the study by Hezaveh et al. (2016) shows that future ALMA data have the potential of constraining the abundance of dark matter subhalos down to $M_{\rm sub} \sim 2 \times 10^7\,\mathrm{M}_\odot$. Independently, Woldesenbet & Williams (2015) have reached similar conclusions (i.e. the necessity of $\gtrsim 10^8\,\mathrm{M}_\odot$ substructures projected near the Einstein radius) by using a model-free analysis of image positions in the relative angle space.

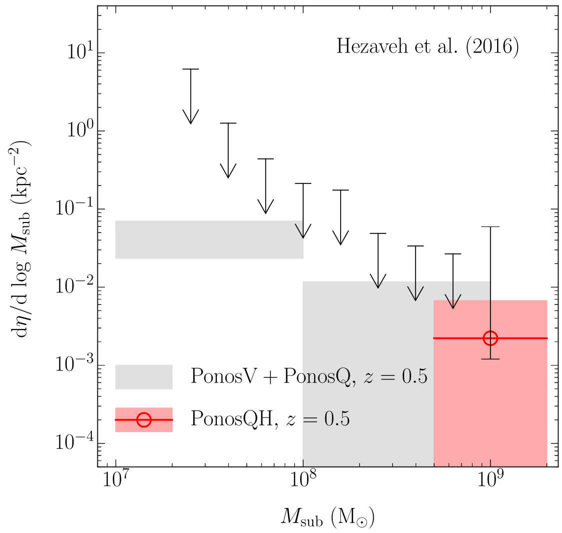

Figure 15. The differential projected mass function of subhalos at the Einstein radius. The errorbars indicate the 95% confidence limits on the projected number density of subhalos around the dusty galaxy SDP.81 (Hezaveh et al. 2016). For comparison, the shaded grey rectangles show the 90% confidence region in a decade of mass at $z = 0.5$ from the combined PonosQ and PonosV simulation sets. The red circle with the red error band shows the mean and 95% confidence limits in PonosQH in the mass bin $0.5 - 2 \times 10^9\,\mathrm{M}_\odot$.

Our hydro simulations do not have enough resolution to generate a realistic substructure population at masses $M_{\rm sub} \lesssim 5 \times 10^8\,\mathrm{M}_\odot$, and higher resolution runs with dissipation are in the making in order to extend our analysis to smaller-mass subhalos. It is also necessary to perform more N-body simulations of early-type galaxy halos in order to obtain unbiased samples and reliable statistics of the subhalo population. We plan to perform a statistical exploration of flux anomalies induced by substructures near the Einstein radius on a larger halo sample. Since our results emphasize the relevance of dwarf-sized systems, such exploration may rely on a large suite of less challenging lower resolution simulations, as long as they can robustly resolve subhalos with masses $\gtrsim 10^8\,\mathrm{M}_\odot$. This is instrumental to broadly asses the impact of substructures on the occurrence of flux anomalies in lensed systems.

We thank the anonymous Referee for useful comments that improved the quality of this work. We thank Lucio Mayer for useful discussions. We thank Oliver Hahn and the AGORA collaboration for help with the initial conditions of the simulations, and Yashar D. Hezaveh for providing the data plotted in Figure 15. Support for this work was provided to P.M. by the NSF through grant AST-1229745, by NASA through grant NNX12AF87G, and by the Pauli Center for Theoretical Studies. D.F. is supported by the Swiss National Science Foundation under grant 200021_140645.

REFERENCES

Agertz, O., Kravtsov, A. V., Leitner, S. N., & Gnedin, N. Y. 2013, ApJ, 770, 25
Agertz, O., Teyssier, R., & Moore, B. 2009, MNRAS, 397, L64
Amara, A., Metcalf, R. B., Cox, T. J., & Ostriker, J. P. 2006, MNRAS, 367, 1367
Auger, M. W., Treu, T., Bolton, A. S., et al. 2010, ApJ, 724, 511
Basu, S., & DasGupta, A. 1997, Theory of Probability & Its Applications, 41, 210
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
Blumenthal, G. R., Faber, S. M., Flores, R., & Primack, J. R. 1986, ApJ, 301, 27
Bode, P., Ostriker, J. P., & Turok, N. 2001, ApJ, 556, 93
Bolton, A. S., Treu, T., Koopmans, L. V. E., et al. 2008, ApJ, 684, 248
Bradač, M., Schneider, P., Lombardi, M., et al. 2004, A&A, 423, 797
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., et al. 2003, MNRAS, 341, 13
Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80
Cappellari, M., Romanowsky, A. J., Brodie, J. P., et al. 2015, ApJL, 804, L21
Cembranos, J. A. R., Feng, J. L., Rajaraman, A., & Takayama, F. 2005, Physical Review Letters, 95, 181301
Chae, K.-H., Bernardi, M., & Kravtsov, A. V. 2014, MNRAS, 437, 3670
Chen, J., Koushiappas, S. M., & Zentner, A. R. 2011, ApJ, 741, 117
Chen, J., Kravtsov, A. V., & Keeton, C. R. 2003, ApJ, 592, 24
Chen, J., Rozo, E., Dalal, N., & Taylor, J. E. 2007, ApJ, 659, 52
Chiba, M. 2002, ApJ, 565, 17
Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25
Dehnen, W. 2000, ApJL, 536, L39
—. 2002, Journal of Computational Physics, 179, 27
Diemand, J., Kuhlen, M., & Madau, P. 2007, ApJ, 657, 262
Diemand, J., Kuhlen, M., Madau, P., et al. 2008, Nature, 454, 735
El-Zant, A., Shlosman, I., & Hoffman, Y. 2001, ApJ, 560, 636
Fakhouri, O., Ma, C.-P., & Boylan-Kolchin, M. 2010, MNRAS, 406, 2267
Feng, J. L. 2010, ARA&A, 48, 495
Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, PASP, 110, 761
Fiacconi, D., Feldmann, R., & Mayer, L. 2015, MNRAS, 446, 1957
Gill, S. P. D., Knebe, A., & Gibson, B. K. 2004, MNRAS, 351, 399
Gnedin, O. Y., Kravtsov, A. V., Klypin, A. A., & Nagai, D. 2004, ApJ, 616, 16
Governato, F., Brook, C., Mayer, L., et al. 2010, Nature, 463, 203
Guedes, J., Callegari, S., Madau, P., & Mayer, L. 2011, ApJ, 742, 76
Haardt, F., & Madau, P. 2012, ApJ, 746, 125
Hahn, O., & Abel, T. 2011, MNRAS, 415, 2101
Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, ArXiv e-prints, arXiv:1601.01388
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581
Hsueh, J.-W., Fassnacht, C. D., Vegetti, S., et al. 2016, ArXiv e-prints, arXiv:1601.01671
Hu, W., Barkana, R., & Gruzinov, A. 2000, Physical Review Letters, 85, 1158
Inoue, K. T. 2016, ArXiv e-prints, arXiv:1601.04414
Inoue, K. T., Minezaki, T., Matsushita, S., & Chiba, M. 2016, MNRAS, 457, 2936
Inoue, K. T., & Takahashi, R. 2012, MNRAS, 426, 2978
Keeton, C. R., Gaudi, B. S., & Petters, A. O. 2003, ApJ, 598, 138
Keeton, C. R., & Moustakas, L. A. 2009, ApJ, 699, 1720
Keller, B. W., Wadsley, J., Benincasa, S. M., & Couchman, H. M. P. 2014, MNRAS, 442, 3013
Kim, C.-G., & Ostriker, E. C. 2015, ApJ, 802, 99
Kim, J.-h., Abel, T., Agertz, O., et al. 2014, ApJS, 210, 14
Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82
Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102
Knollmann, S. R., & Knebe, A. 2009, ApJS, 182, 608
Kochanek, C. S., Falco, E. E., Impey, C. D., et al. 2000, ApJ, 543, 131
Koopmans, L. V. E. 2005, MNRAS, 363, 1136
Kravtsov, A., Vikhlinin, A., & Meshscheryakov, A. 2014, ArXiv e-prints, arXiv:1401.7329
Kroupa, P. 2001, MNRAS, 322, 231
Li, R., Frenk, C. S., Cole, S., et al. 2015, ArXiv e-prints, arXiv:1512.06507
Macciò, A. V., Moore, B., Stadel, J., & Diemand, J. 2006, MNRAS, 366, 1529
Madau, P., Shen, S., & Governato, F. 2014, ApJL, 789, L17
Mao, S., Jing, Y., Ostriker, J. P., & Weller, J. 2004, ApJL, 604, L5
Mao, S., & Schneider, P. 1998, MNRAS, 295, 587
McKean, J. P., Koopmans, L. V. E., Flack, C. E., et al. 2007, MNRAS, 378, 109
Metcalf, R. B. 2005, ApJ, 629, 673
Metcalf, R. B., & Amara, A. 2012, MNRAS, 419, 3414
Metcalf, R. B., & Madau, P. 2001, ApJ, 563, 9
Metcalf, R. B., & Zhao, H. 2002, ApJL, 567, L5
Moore, B., Ghigna, S., Governato, F., et al. 1999, ApJL, 524, L19
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Pillepich, A., Kuhlen, M., Guedes, J., & Madau, P. 2014, ApJ, 784, 161
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, ArXiv e-prints, arXiv:1502.01589
Pontzen, A., & Governato, F. 2012, MNRAS, 421, 3464
—. 2014, Nature, 506, 171
Power, C., Navarro, J. F., Jenkins, A., et al. 2003, MNRAS, 338, 14
Roškar, R., Fiacconi, D., Mayer, L., et al. 2015, MNRAS, 449, 494
Serra, P., Oosterloo, T., Morganti, R., et al. 2012, MNRAS, 422, 1835
Shen, S., Madau, P., Conroy, C., Governato, F., & Mayer, L. 2014, ApJ, 792, 99
Shen, S., Madau, P., Guedes, J., et al. 2013, ApJ, 765, 89
Shen, S., Wadsley, J., & Stinson, G. 2010, MNRAS, 407, 1581
Spergel, D. N., & Steinhardt, P. J. 2000, Physical Review Letters, 84, 3760
Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685
Stadel, J. 2013, PkdGRAV2: Parallel fast-multipole cosmological code, Astrophysics Source Code Library, ascl:1305.005
Stadel, J., Potter, D., Moore, B., et al. 2009, MNRAS, 398, L21
Stinson, G., Seth, A., Katz, N., et al. 2006, MNRAS, 373, 1074
Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T., & Bolton, A. S. 2014, MNRAS, 442, 2017
Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341
Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, Phys. Rev. D, 88, 043502
Wadsley, J. W., Stadel, J., & Quinn, T. 2004, Nature, 9, 137
Wambsganss, J., Bode, P., & Ostriker, J. P. 2005, ApJL, 635, L1
Weinberg, D. H., Bullock, J. S., Governato, F., Kuzio de Naray, R., & Peter, A. H. G. 2015, Proceedings of the National Academy of Science, 112, 12249
Wilman, D. J., & Erwin, P. 2012, ApJ, 746, 160
Woldesenbet, A. G., & Williams, L. L. R. 2015, MNRAS, 454, 862
Xu, D., Sluse, D., Gao, L., et al. 2015, MNRAS, 447, 3189
Xu, D. D., Mao, S., Wang, J., et al. 2009, MNRAS, 398, 1235
Zemp, M., Stadel, J., Moore, B., & Carollo, C. M. 2007, MNRAS, 376, 273
Zentner, A. R., Berlind, A. A., Bullock, J. S., Kravtsov, A. V., & Wechsler, R. H. 2005, ApJ, 624, 505
Zhu, Q., Marinacci, F., Maji, M., et al. 2015, ArXiv e-prints, arXiv:1506.05537

Preprint typeset using LaTeX style emulateapj v. 01/23/15

SWIFT COALESCENCE OF SUPERMASSIVE BLACK HOLES IN COSMOLOGICAL MERGERS OF MASSIVE GALAXIES

Fazeel Mahmood Khan
Department of Space Science, Institute of Space Technology, PO Box 2750 Islamabad, Pakistan

Davide Fiacconi, Lucio Mayer
Center for Theoretical Astrophysics and Cosmology, Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Peter Berczik[1,2]
National Astronomical Observatories and Key Laboratory of Computational Astrophysics, Chinese Academy of Sciences, 20A Datun Rd., Chaoyang District, 100012, Beijing, China

and

Andreas Just
Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstrasse 12-14, 69120, Heidelberg, Germany

ABSTRACT

Supermassive black holes (SMBHs) are ubiquitous in galaxies with a sizable mass. It is expected that a pair of SMBHs originally in the nuclei of two merging galaxies would form a binary and eventually coalesce via a burst of gravitational waves. So far theoretical models and simulations have been unable to predict directly the SMBH merger timescale from ab-initio galaxy formation theory, focusing only on limited phases of the orbital decay of SMBHs under idealized conditions of the galaxy hosts. The predicted SMBH merger timescales are long, of order Gyrs, which could be problematic for future gravitational wave searches. Here we present the first multi-scale ΛCDM cosmological simulation that follows the orbital decay of a pair of SMBHs in a merger of two typical massive galaxies at $z \sim 3$, all the way to the final coalescence driven by gravitational wave emission. The two SMBHs, with masses $\sim 10^8$ $M_\odot$, settle quickly in the nucleus of the merger remnant. The remnant is triaxial and extremely dense due to the dissipative nature of the merger and the intrinsic compactness of galaxies at high redshift. Such properties naturally allow a very efficient hardening of the SMBH binary. The SMBH merger occurs in only ~ 10 Myr after the galactic cores have merged, which is two orders of magnitude smaller than the Hubble time.

Keywords: black hole physics – galaxies: interactions – galaxies: kinematics and dynamics – galaxies: nuclei – gravitational waves – methods: numerical

1. INTRODUCTION

Dual active galactic nuclei (AGNs) at kiloparsec to hundred parsec separations have been detected (Comerford et al. 2013), but at smaller separations there are only unconfirmed candidates (Eracleous et al. 2012; Graham et al. 2015). At the same time, the orbital decay of two supermassive black holes (SMBHs) at the center of merging galaxies (Begelman et al. 1980) has been theoretically studied with a variety of computer simulations, but always neglecting one or more important processes. Substantial work has focused on the gravitational interaction between the SMBHs and the stellar background, as it would be appropriate for a gas-free galaxy (Makino & Funato 2004; Vasiliev et al. 2015), as well as on the dynamics of two SMBHs within a dissipative gaseous background, though neglecting gravitational encounters with individual stars (Dotti et al. 2007; Mayer et al. 2007; Chapon et al. 2013). Furthermore, individual simulations, due to computational limitations, typically follow only a limited phase of the SMBH pair evolution, either before or after a Keplerian binary forms (Mayer 2013). It has been noticed that SMBHs in stellar systems may stall at parsec separations as the loss cone is depleted, rendering the transfer of energy between the binary and the stellar background inefficient (Makino & Funato 2004; Berczik et al. 2005). This has been dubbed the "last parsec problem" (Milosavljević & Merritt 2001). However, it has been shown that the loss cone can be refilled if the potential of the galaxy has substantial deviations from sphericity (Khan et al. 2011; Preto et al. 2011; Vasiliev et al. 2015). Yet, extrapolating the hardening rates seen in these recent simulations to the gravitational wave (GW) dominated phase using analytical models, the resulting SMBH merger timescales are ~ 1 Gyr (Khan et al. 2012a,b). The galaxy merger timescale is also of order a Gyr (Stewart et al. 2009), suggesting that the overall process takes a significant fraction of the age of the Universe. At $z > 2$, when the lookback time is also of order a few Gyr, this would im-

---

khanfazeel.ist@gmail.com

[1] Main Astronomical Observatory, National Academy of Sciences of Ukraine, 27 Akademika Zabolotnoho St., 03680, Kyiv, Ukraine

[2] Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstrasse 12-14, 69120, Heidelberg, Germany

ply low SMBH coalescence rates, a potential problem for future GW experiments such as the Evolved Laser Interferometer Space Antenna (eLISA) (Amaro-Seoane et al. 2013).

If gas is present, such as in circumnuclear disks forming as a result of gas-rich mergers, the orbital decay of the SMBHs proceeds on a faster track, leading to a hard binary with pc-scale separation in $\sim 1-100$ Myr, depending on the clumpiness of the interstellar medium (Mayer et al. 2007; Chapon et al. 2013; Fiacconi et al. 2013; Roškar et al. 2015). However, the drag may become inefficient at smaller separations, potentially resulting in a stalling binary (Chapon et al. 2013; Mayer 2013).

The results of all these simulations depend strongly on the mass distribution and properties of the stars and interstellar gas of the circumnuclear region, which are inherited from idealized initial conditions. Therefore it is still unclear how and at what pace the orbital decay of SMBHs proceeds in realistic galaxy mergers. In order to make progress, in this Letter we have carried out the first ab-initio calculation of a SMBH merger that starts from a galaxy merger pinpointed in a state-of-the-art cosmological hydrodynamical simulation. We follow the SMBHs all the way to the final spiral-in phase with the aid of post-Newtonian corrections (Blanchet 2006).

2. GALAXY MERGER SIMULATION

We identify the merger of two massive galaxies at $z \sim 3.5$ in the Argo cosmological hydrodynamical simulation (Feldmann & Mayer 2015; Fiacconi et al. 2015). The simulation follows the formation of a group-sized halo with mass $\approx 2 \times 10^{13}$ $\mathrm{M}_\odot$ at $z = 0$, and includes gas cooling, star formation (SF) and a supernovae feedback models that have been shown to produce realistic galaxies at a variety of mass scales (Governato et al. 2010; Guedes et al. 2011). The halo evolves in a mildly over-dense region and its virial mass is close to the characteristic scale $M^\star$ of the halo mass function at low z, suggesting that it should be a common host for massive galaxies (Feldmann & Mayer 2015; Fiacconi et al. 2015). In lower resolution simulations the group hosts a central galaxy with properties typical of massive early-types at $z = 0$ (Feldmann et al. 2010).

The central galaxy of the Argo simulation undergoes its last major merger (with a stellar mass ratio ~ 0.3) at $z \simeq 3.5$. The merger involves two disk-like galaxies in a nearly parabolic (slightly hyperbolic) orbit, with their stellar spins misaligned by $\approx 67^\circ$. Such a configuration is typical for major mergers in ΛCDM cosmology (Khochfar & Burkert 2006). The two galaxies have stellar masses $M_{\star,1} \approx 3.6 \times 10^{10}$ $\mathrm{M}_\odot$ and $M_{\star,2} \approx 10^{10}$ $\mathrm{M}_\odot$, and gas fractions $f_1 \approx 7.7\%$ and $f_2 = 11.5\%$, respectively. The two galaxies and their group environment are shown in Figure 1 when they are about to merge.

The cosmological simulation does not originally contain any SMBH, and its resolution would not allow us to probe the evolution of a BH binary. Therefore, we increase the resolution by performing static particle splitting (Mayer et al. 2007; Roškar et al. 2015). Specifically, we extract from the cosmological simulation a spherical region with radius ~ 13.5 kpc at $z = 3.6$ that encompasses the two galaxies and part of their environment. We check that the average dynamical time of such region is $\gtrsim 100$ Myr, which is larger than the dynamical time in the central region and the simulation time that we target, about a few tens of Myr. At this stage, the cores of the two galaxies are at a separation of $\lesssim 4$ kpc. We then split all particle species in 8 child particles with masses 8 times smaller and the same velocity of the parent particle, thus conserving mass and linear momentum exactly, and angular momentum at the kernel level. Thermodynamic properties of gas particles (i.e. density and temperature) are interpolated among the child particles (Roškar et al. 2015), while child stellar particles maintain the properties of their parents (e.g. the age). After the splitting, the simulation contains 9 452 581 stellar particles, 1 088 920 gas particles, and 1 669 922 dark matter particles with masses 6.4×10^3 $\mathrm{M}_\odot$, 2×10^4 $\mathrm{M}_\odot$, and 10^5 $\mathrm{M}_\odot$, respectively. We reduce to $\epsilon = 5$ pc the gravitational softenings of gaseous and stellar particles to increase the spatial resolution, while the dark matter softening is reduced by a factor $8^{1/3} = 2$ to maintain the local density and the smooth gravitational field. We extensively tested this procedure by running twin simulations with $\epsilon = 15$ and 50 pc; we checked that (i) no spurious effects on scales larger than the original softening were introduced, and (ii) the dynamics of the introduced SMBHs (see below) converged down to the adopted softenings.

During the splitting procedure, we introduce two SMBHs at the local minima of the gravitational potential of the galactic cores. We assign to the SMBHs the mass-weighted average velocity of all the particles within 250 pc from their positions. We checked that the velocities do not depend strongly on the size of the regions that we choose. Since the two galaxies are relatively gas-poor at $z \sim 3.5$ and we aim to continue the simulation for $\sim 20-40$ Myr, we neglect mass accretion and feedback from the SMBHs, that are thus treated as collisionless particles. The SMBHs have the same softening as stellar and gas particles. We finally choose their masses according to local scaling relations (Ferrarese & Merritt 2000; Tremaine et al. 2002; McConnell & Ma 2013). Specifically, we use the latest determination of the BH mass-velocity dispersion relation (McConnell & Ma 2013) and we determine the SMBH masses after measuring the average velocity dispersion within one half-mass radius for each galaxy before performing particle splitting. The resulting masses are $M_{\bullet,1} = 3 \times 10^8$ $\mathrm{M}_\odot$ and $M_{\bullet,2} = 8 \times 10^7$ $\mathrm{M}_\odot$, with a mass ratio $M_{\bullet,1}/M_{\bullet,2} = 3.75$. Though using the local scaling relations for the SMBH masses is formally inappropriate for high-z galaxies, this might result to be a conservative choice because there are both observational (Merloni et al. 2010; Trakhtenbrot et al. 2015) and theoretical (DeGraf et al. 2015) hints that the normalization of those relations increases at high-z, i.e. galaxies host proportionally larger SMBHs.

After setting-up the initial conditions as described, we simulate the final stages of the galaxy merger until the separation of the two SMBHs is about the resolution. We use the GASOLINE code (Wadsley et al. 2004), but including additional sub-resolution physics. Specifically, we add the gas radiative cooling from metal lines, a pressure floor to avoid spurious fragmentation, and an equilibrium temperature-density relation for gas denser than 0.1 H cm^{-3} to model the optically-thick phase, calibrated on 2D radiative transfer simulations in typical starburst conditions (Spaans & Silk 2000; Roškar et al.

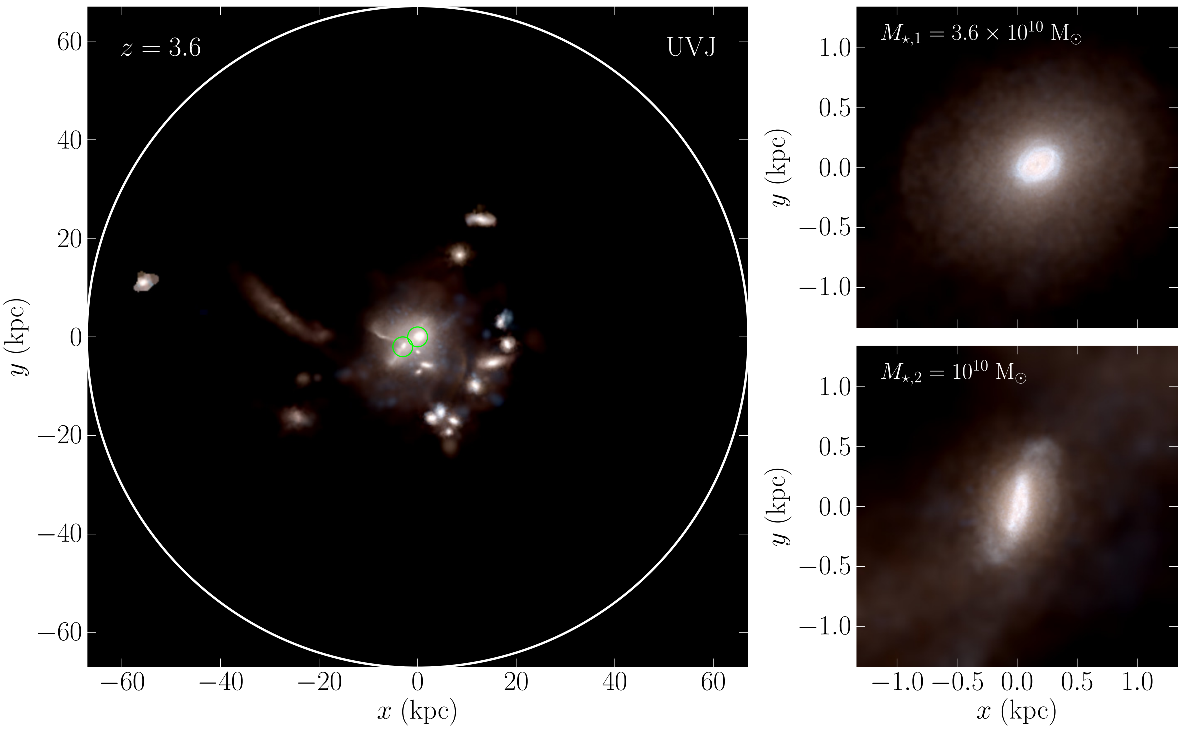

Figure 1. Group environment of the galaxy merger. The left panel shows a mock UVJ map of the galaxy group at $z = 3.6$. The white circle marks the virial radius of the group halo, while the green circles mark the merging galaxies. The upper-right and lower-right panels show a zoom-in on the central galaxy of the group and the interacting companion, respectively. Lengths are in physical coordinates.

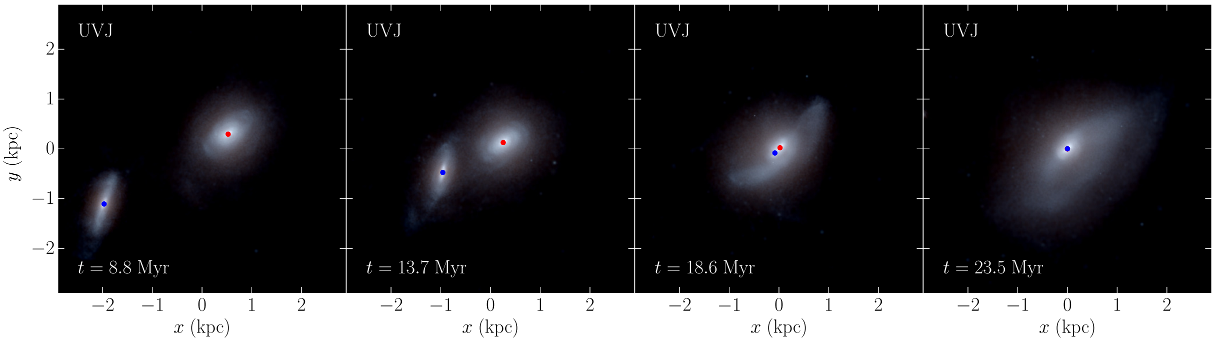

Figure 2. From left to right: time evolution of the galaxy merger after the beginning of the re-sampled, higher-resolution simulation. Each panel shows a mock UVJ photometric image of the merger, and the red and blue dots mark the position of the primary and secondary BH, respectively. Lengths are in physical coordinates.

2015). We also increase the density threshold to form star to 1000 H cm^{-3} and we reduce the temperature threshold to 300 K, to account for the new cooling.

Figure 2 shows different stages of the merger of the two galaxies in our simulation after we perform the particle splitting. The Figure reveals that the two galaxies are flattened and disk-like. The remnant has an elongated shape out to a few kpc soon after the merger.

3. SUPERMASSIVE BLACK HOLE BINARY EVOLUTION

The left panel of Figure 3 describes the orbital evolution of the two SMBHs starting form ~ 4 kpc till the final coalescence. As the two galaxies merge, their SMBHs sink in the remnant surrounded by stellar cusps bound to them. The orbital decay is governed by dynamical friction of the stellar cusps against the stellar, gas and dark matter background originating from the merger of the two hosts.

During the final stage of the merger (i.e. at $t \approx 20$ Myr after the particle splitting) the merger remnant is gas poor (gas fraction $\sim 5\%$) owing to gas consumption by SF. Stars dominate the enclosed mass out to ~ 3 kpc and provide the dominant contribution to the dynamical friction exerted by the background. Figure 4 shows the mass distribution of the individual components when the

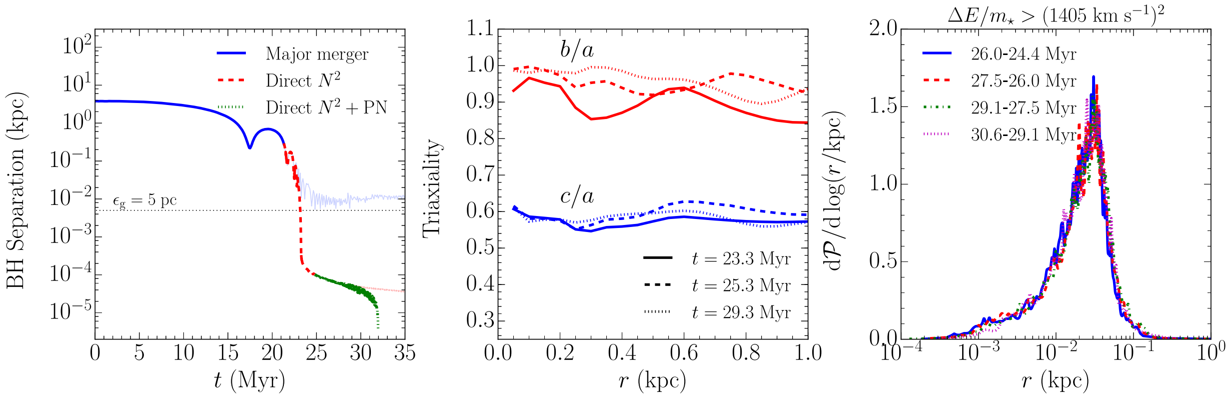

Figure 3. Left panel: time evolution of the separation between the SMBHs. Blue-solid, red-dashed, and green-dotted lines show the evolution during the hydrodynamical, re-sampled simulation of the merger, the direct N-body calculation, and after having introduced post-Newtonian corrections, respectively. Thin and light versions of the same lines refer to the continuation of the respective simulations. The horizontal dotted line marks the gravitational softening of the hydrodynamical simulation. Central panel: radial profiles of the ratio b/a (red) and c/a (blue) between the principal axes of the moment of inertia tensor ($c \leq b \leq a$) at different times: 23.3 Myr (solid), 25.3 Myr (dashed), and 29.3 Myr (dotted). Right panel: probability density function of the radial distance from the center of the merger remnant for the stellar particles that have interacted with the central binary across 26-24.4 Myr (blue, solid), 27.5-26 Myr (red, dashed), 29.1-27.5 Myr (green, dot-dashed), and 30.6-29.1 Myr (magenta, dotted).

separation of the two SMBHs reaches about 300 pc, i.e. ≈ 21.5 Myr after the particle splitting. The stellar mass is almost 2 orders of magnitude larger than the gas one over all spatial scales except in the central 10 pc, where the difference is about a factor of 20.

Then, we extract a spherical region of 5 kpc at $t \sim 21.5$ Myr after particle splitting around the more massive SMBH to initialize a direct N-body simulation containing in total $\sim 6 \times 10^6$ particles. We treat the remaining gas particles in the selected volume as stars, since they are sub-dominant in mass. Almost the entire stellar mass is within 5 kpc, so an artificial cut-off at 5 kpc shall not introduce significant changes in stellar mass profile in the inner region for follow up evolution. However, at truncation separation, the dark matter has a steeply rising mass profile; we compare it with a later snapshot during the N-body evolution at $t \approx 30$ Myr. We do not observe a noticeable evolution from outer to inner region.

We further evolve the selected region using the high-performance ϕ-GPU code (Berczik et al. 2011). At the end of our previous galaxy merger simulation, stellar and gas particles have a softening of 5 pc and dark matter particles have a softening of 150 pc. We start our direct N-body run by decreasing the stellar softening to 0.1 pc while keeping the dark matter particles softening unchanged to avoid two-body relaxation effects as the latter have a relatively large mass. Figure 4 shows that the mass of the dark matter component does not increase in the central region during the whole evolution period. The softening parameter for the force calculation between the two black holes is set to 0. In order to calculate the softening between different particle species we employ the following criterion:

$$\epsilon_{ij}^2 = (\epsilon_i^2 + \epsilon_j^2)/2, \tag{1}$$

where $\epsilon_\bullet = 0$ for both black holes, $\epsilon_\star = 0.1$ pc for stars and $\epsilon_{\rm dm} = 125$ pc for dark matter particles. In star-black hole interactions we further reduce the softening to 0.007 pc, which is smaller than the semi-major axis of the binary when the gravitational wave emission dominates (Figure 3, left panel). In order to take into account energy loss by gravitational wave emission, we incorporate post-newtonian terms up to 3.5 in the equation of motion of the binary SMBHs (Blanchet 2006).

Dynamical friction efficiently shrinks the separation between the two SMBHs and they form a binary once individual cusps merge at $t \sim 23.5$ Myr. The separation drops rapidly to ~ 0.3 pc in less than 1 Myr, owing to the high nuclear density, until the binary gets hard and dynamical friction becomes inefficient. The subsequent phase of the decay is dominated by three-body encounters between the binary and the surrounding stars. This phase is the longest, taking ≈ 8 Myr before the separation decreases to ~ 0.01 pc, at which point GW emission takes over and brings the SMBHs to rapid coalescence in 2 Myr (Figure 3). Figure 3 also shows that post-Newtonian terms are crucial for the sinking of the binary already at a separation of $\gtrsim 0.03$ pc. The orbital decay rate in the post-Newtonian phase is in rough agreement with simple semi-analytical predictions based on orbit-averaged expressions (Peters & Mathews 1963), which do not take into account the contributions from higher order terms.

We define the coalescence time, and stop the simulations, when the separation is $< 4(r_{\rm s,1}+r_{\rm s,2})$, where $r_{{\rm s},j}$ is the Schwarzschild radius of the j-th BH, as following the evolution further would require a fully relativistic treatment. The coalescence, counting from the merger of the two cusps at $t \sim 23.5$ Myr, takes less than 10 Myr, which is roughly two orders of magnitude faster than previous decay time estimates inferred for non-cosmological, purely stellar hosts (when rescaled to nearby galaxies) (Khan et al. 2012b, 2013). The preceding large-scale approach and merger of the two galaxies lasts ~ 200 Myr, so that the overall process is completed in significantly less than the lookback time at $z \sim 3.5$.

The shape of the merger remnant over time is shown in the central panel of Figure 3. It was obtained by measuring the moments of inertia tensor of a homogeneous ellipsoid. The remnant is clearly triaxial at all times. The right panel of Figure 3 shows the distribution of the mean radial distances of the stars that contribute to the change

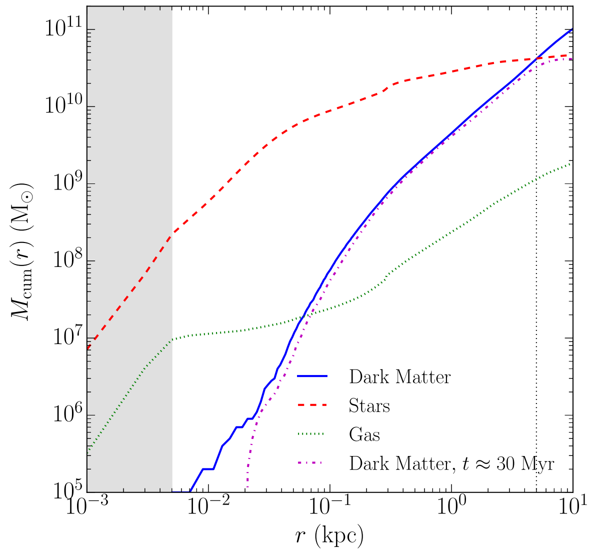

Figure 4. Enclosed mass profile of dark matter (blue, continuous), stars (red, dashed) and gas (green, dotted) at $t \sim 21.5$ Myr when we select the inner 5 kpc (vertical dotted line) for the direct N-body simulation. The gray area marks $\epsilon = 5$ pc. The magenta dot-dashed line shows the dark matter profile at $t \approx 30$ Myr, which is not modified by our truncation.

in the binding energy of the binary at different times. Those have been identified statistically as the stars that undergo large specific total energy change between two subsequent snapshots, $\Delta E/m_\star > (1405 \text{ km s}^{-1})^2$. This absolute threshold depends on the binary properties only, $\Delta E/m_\star \approx G\ \mu_\bullet\ C\ a^{-1}$, where $\mu_\bullet$ is the binary reduced mass, a is the binary separation, and $C \approx 2$, as inferred from three-body encounter simulations (Hills 1983; Quinlan 1996). This criterion allows us to select only the stars involved in encounters with the SMBH binary, because the energy changes due to the large scale evolution of the system and to two-body encounters with other stars and dark matter particles are small compared to the adopted threshold. Most of those stars come from 10-100 pc, which is at quite far from the BH binary sitting at the center of the remnant. This shows that the loss cone is efficiently refilled from stars on plunging orbits.

Finally, we stress that, at the resolution that we are employing, our results on the binary evolution during the phase dominated by stellar encounters are robust. Indeed, previous tests in flat, rotating, and triaxial systems have shown that the hardening rate is almost independent of the number of particles when above $\sim 10^6$ (Berczik et al. 2006; Khan et al. 2011; Preto et al. 2011; Khan et al. 2013; Holley-Bockelmann & Khan 2015). Since the star particles in our run are $\sim 5.5 \times 10^6$, we expect that we reach a regime of statistical convergence where fewer encounters with more massive particles produce a total energy exchange comparable to that occurring with many more encounters with lighter particles.

4. DISCUSSION AND CONCLUSIONS

In this Letter, we discuss the first multi-scale simulation that probes the evolution of a SMBH binary forming within a cosmological major merger, all the way down to the coalescence driven by the emission of GWs. We start from the Argo cosmological simulation, where we identify and re-simulate at higher resolution a major merger between two massive galaxies at $z \approx 3.5$. Gas dissipation before the merger is instrumental in creating the conditions for the rapid orbital decay of the SMBHs, which is our key finding. Indeed, the high central stellar density in the remnant is the result of gas inflows in the inner $\lesssim 500$ pc due to prior cosmological gas accretion and mergers (Feldmann et al. 2010). These effects cannot be accounted for in idealized galaxy mergers, rather require cosmological simulations as those employed here. The dense remnant not only causes strong dynamical friction by both gas and stars in the early decay phase, but provides an abundant reservoir of stars in the nuclear region that interact with the binary in the late phases of the decay.

At the same time, triaxiality is necessary to avoid the loss cone problem and is a natural result of mergers between non-spherical galaxy progenitors. Since massive galaxies comprise a large fraction of star forming galaxies at high z, many with massive disk-like components (Wisnioski et al. 2015), the pre-merger conditions in our simulation should be typical of the progenitors of massive early-type galaxies that host the most massive SMBHs at low z.

We have presented only one multi-scale simulation due to the high computational cost that such calculations entail. In order to understand how general is this result we revisited coalescence times obtained by a large suite of N-body simulations in Khan et al. (2012b). We rescale[3] their merger product with the one obtained in our cosmological study. We find SMBH merger timescales from the time of binary formation in the range $10-30$ Myr, which satisfactorily match our results. This suggests that the timescale is primarily determined by the characteristic density of the host in the nucleus. Since the stellar mass, $M_\star \lesssim 10^{11}$ $M_\odot$, and the effective radius, $r_{\rm eff} \approx 600$ pc, of our galaxy are typical of observed $z \gtrsim 2$ massive early-type galaxies (Bezanson et al. 2009; Szomoru et al. 2012; Bezanson et al. 2013), we conclude that SMBH mergers in those systems at $z > 2$ should be generically as fast as we find here.

In turn, the scaling argument suggests that the much longer coalescence timescales, of order a Gyr, should be the norm for massive, less dense early-type galaxies at low redshift. These are the galaxies hosting SMBHs with masses $> 10^8$ $M_\odot$, whose mergers should be detectable by Pulsar Timing Arrays (PTAs; Hobbs et al. 2010). Recently, it has been argued that the lack of detection by PTAs might be difficult to reconcile with simple analytical predictions of the hardening rate which assume full loss cone and yield short SMBH merging timescales $\lesssim 10^8$ yr (Shannon et al. 2015). However, here we argue that such short coalescence timescales would occur only at high redshift, hence outside the observability window of PTAs. At low redshift, coalescence times are of order of Gyr; the preceding galaxy merger phase up to SMBH binary formation will also be delayed as major mergers become more rare, with less than 1 merger per galaxy per

---

[3] This rescaling is possible because the models by Khan et al. (2012b) are scale-free and physical scaling can be obtained by comparing some characteristic length and mass of the model, e.g. influence radius and mass of SMBH to some reference values.

Gyr expected below $z = 1$ for $L > L^\star$ galaxies (Stewart et al. 2009). Such a low merger rate would likely yield a GW background signal below the detection limit of PTAs, naturally explaining the current lack of detection.

Nevertheless, early-type galaxies with properties expected for a recent dry merger (e.g. shells, tidal tails, and little recent SF) could be the ideal target for detecting SMBH binaries. The fast coalescence times that we find support optimistic expectations for the number of GW emission events detectable with eLISA, at least for the most massive SMBHs in its detection window ($z \sim 2-6$), in the range $10^6 - 10^8$ M$_\odot$. The forecasts assume nearly instantaneous SMBH mergers after the galaxies merge (Amaro-Seoane et al. 2013). Indeed, based on the known scaling relations, the host galaxies of such black holes should have stellar masses above $10^9 - 10^{11}$ M$_\odot$, the mass range at which SF is most efficient in galaxies, yielding dense, triaxial merger remnants that would assist the prompt coalescence of their SMBHs.

We thank Robert Feldmann for providing us with the Argo simulation snapshots. We thank Pedro R. Capelo and Alberto Sesana for useful discussions and for a thorough reading of the manuscript. We thank Rainer Spurzem for support through the Silk Road Project at National Astronomical Observatories of Chinese Academy of Sciences. F.M.K. acknowledges support by the Excellenzinitiative II "Mobilitätsmaßnahmen im Rahmen internationaler Forschungskooperationen 2015-16" of Heidelberg University. D.F. is supported by the Swiss National Science Foundation under grant #No. 200021_140645. P.B. is supported by the Strategic Priority Research Program "The Emergence of Cosmological Structure" of the Chinese Academy of Sciences (No. XDB09000000), the Sonderforschungsbereich SFB 881 "The Milky Way System" (subproject Z2) of the German Research Foundation (DFG)and by the NASU under the Main Astronomical Observatory GRID/GPU computing cluster project.

REFERENCES

Amaro-Seoane, P., Aoudia, S., Babak, S., et al. 2013, GW Notes, Vol. 6, p. 4-110, 6, 4

Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, Nature, 287, 307

Berczik, P., Merritt, D., & Spurzem, R. 2005, ApJ, 633, 680

Berczik, P., Merritt, D., Spurzem, R., & Bischof, H.-P. 2006, ApJL, 642, L21

Berczik, P., Nitadori, K., Zhong, S., et al. 2011, in International conference on High Performance Computing, Kyiv, Ukraine, October 8-10, 2011., p. 8-18, 8–18

Bezanson, R., van Dokkum, P. G., Tal, T., et al. 2009, ApJ, 697, 1290

Bezanson, R., van Dokkum, P. G., van de Sande, J., et al. 2013, ApJL, 779, L21

Blanchet, L. 2006, Living Reviews in Relativity, 9, 4

Chapon, D., Mayer, L., & Teyssier, R. 2013, MNRAS, 429, 3114

Comerford, J. M., Schluns, K., Greene, J. E., & Cool, R. J. 2013, ApJ, 777, 64

DeGraf, C., Di Matteo, T., Treu, T., et al. 2015, MNRAS, 454, 913

Dotti, M., Colpi, M., Haardt, F., & Mayer, L. 2007, MNRAS, 379, 956

Eracleous, M., Boroson, T. A., Halpern, J. P., & Liu, J. 2012, ApJS, 201, 23

Feldmann, R., Carollo, C. M., Mayer, L., et al. 2010, ApJ, 709, 218

Feldmann, R., & Mayer, L. 2015, MNRAS, 446, 1939

Ferrarese, L., & Merritt, D. 2000, ApJL, 539, L9

Fiacconi, D., Feldmann, R., & Mayer, L. 2015, MNRAS, 446, 1957

Fiacconi, D., Mayer, L., Roškar, R., & Colpi, M. 2013, ApJL, 777, L14

Governato, F., Brook, C., Mayer, L., et al. 2010, Nature, 463, 203

Graham, M. J., Djorgovski, S. G., Stern, D., et al. 2015, Nature, 518, 74

Guedes, J., Callegari, S., Madau, P., & Mayer, L. 2011, ApJ, 742, 76

Hills, J. G. 1983, AJ, 88, 1269

Hobbs, G., Archibald, A., Arzoumanian, Z., et al. 2010, Classical and Quantum Gravity, 27, 084013

Holley-Bockelmann, K., & Khan, F. M. 2015, ApJ, 810, 139

Khan, F. M., Berentzen, I., Berczik, P., et al. 2012a, ApJ, 756, 30

Khan, F. M., Holley-Bockelmann, K., Berczik, P., & Just, A. 2013, ApJ, 773, 100

Khan, F. M., Just, A., & Merritt, D. 2011, ApJ, 732, 89

Khan, F. M., Preto, M., Berczik, P., et al. 2012b, ApJ, 749, 147

Khochfar, S., & Burkert, A. 2006, A&A, 445, 403

Makino, J., & Funato, Y. 2004, ApJ, 602, 93

Mayer, L. 2013, Classical and Quantum Gravity, 30, 244008

Mayer, L., Kazantzidis, S., Madau, P., et al. 2007, Science, 316, 1874

McConnell, N. J., & Ma, C.-P. 2013, ApJ, 764, 184

Merloni, A., Bongiorno, A., Bolzonella, M., et al. 2010, ApJ, 708, 137

Milosavljević, M., & Merritt, D. 2001, ApJ, 563, 34

Peters, P. C., & Mathews, J. 1963, Physical Review, 131, 435

Preto, M., Berentzen, I., Berczik, P., & Spurzem, R. 2011, ApJL, 732, L26

Quinlan, G. D. 1996, New A, 1, 35

Roškar, R., Fiacconi, D., Mayer, L., et al. 2015, MNRAS, 449, 494

Shannon, R. M., Ravi, V., Lentati, L. T., et al. 2015, Science, 349, 1522

Spaans, M., & Silk, J. 2000, ApJ, 538, 115

Stewart, K. R., Bullock, J. S., Barton, E. J., & Wechsler, R. H. 2009, ApJ, 702, 1005

Szomoru, D., Franx, M., & van Dokkum, P. G. 2012, ApJ, 749, 121

Trakhtenbrot, B., Urry, C. M., Civano, F., et al. 2015, Science, 349, 168

Tremaine, S., Gebhardt, K., Bender, R., et al. 2002, ApJ, 574, 740

Vasiliev, E., Antonini, F., & Merritt, D. 2015, ApJ, 810, 49

Wadsley, J. W., Stadel, J., & Quinn, T. 2004, New A, 9, 137

Wisnioski, E., Förster Schreiber, N. M., Wuyts, S., et al. 2015, ApJ, 799, 209

Bibliography

Abadi M. G., Moore B., Bower R. G., 1999, MNRAS, 308, 947

Abbott B. P. et al., 2016, Physical Review Letters, 116, 061102

Abel T., Bryan G. L., Norman M. L., 2002, Science, 295, 93

Agertz O., Kravtsov A. V., Leitner S. N., Gnedin N. Y., 2013, ApJ, 770, 25

Agertz O., Lake G., Teyssier R., Moore B., Mayer L., Romeo A. B., 2009, MNRAS, 392, 294

Agertz O. et al., 2007, MNRAS, 380, 963

Agertz O., Teyssier R., Moore B., 2009, MNRAS, 397, L64

Agol E., Krolik J., Turner N. J., Stone J. M., 2001, ApJ, 558, 543

Aguilar L. A., White S. D. M., 1985, ApJ, 295, 374

Alexander T., Natarajan P., 2014, Science, 345, 1330

Amaro-Seoane P. et al., 2013, GW Notes, Vol. 6, p. 4-110, 6, 4

Antonucci R., 1993, ARA&A, 31, 473

Asida S. M., 2000, ApJ, 528, 896

Baker J. G., Centrella J., Choi D.-I., Koppitz M., van Meter J., 2006, Phys. Rev. D, 73, 104002

Balbus S. A., 2009, MNRAS, 395, 2056

Balbus S. A., Bonart J., Latter H. N., Weiss N. O., 2009, MNRAS, 400, 176

Balbus S. A., Hawley J. F., 1991, ApJ, 376, 214

Balbus S. A., Latter H., Weiss N., 2012, MNRAS, 420, 2457

Balbus S. A., Latter H. N., 2010, MNRAS, 407, 2565

Balbus S. A., Schaan E., 2012, MNRAS, 426, 1546

Balbus S. A., Weiss N. O., 2010, MNRAS, 404, 1263

Ball W. H., Tout C. A., Żytkow A. N., 2012, MNRAS, 421, 2713

Ball W. H., Tout C. A., Żytkow A. N., Eldridge J. J., 2011, MNRAS, 414, 2751

Ballot J., Brun A. S., Turck-Chièze S., 2007, ApJ, 669, 1190

Balsara D. S., 1995, Journal of Computational Physics, 121, 357

Barnes J., Hut P., 1986, Nature, 324, 446

Barnes J. E., 1988, ApJ, 331, 699

Barnes J. E., 1992, ApJ, 393, 484

Barnes J. E., Hernquist L., 1996, ApJ, 471, 115

Baumgarte T. W., Shapiro S. L., 1999, ApJ, 526, 941

Beckmann V., Shrader C. R., 2012, Active Galactic Nuclei

Begelman M. C., 1978, MNRAS, 184, 53

Begelman M. C., 1979, MNRAS, 187, 237

Begelman M. C., 2010, MNRAS, 402, 673

Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307

Begelman M. C., Rossi E. M., Armitage P. J., 2008, MNRAS, 387, 1649

Begelman M. C., Shlosman I., 2009, ApJ, 702, L5

Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

Beletsky Y., Gadotti D. A., Moiseev A., Alves J., Kniazev A., 2011, MNRAS, 418, L6

Bennett A. S., 1962, MmRAS, 68, 163

Bentley J. L., 1979, IEEE Transactions on Software Engineering, SE-5, 333

Bentz M. C., Peterson B. M., Pogge R. W., Vestergaard M., Onken C. A., 2006, ApJ, 644, 133

Berczik P., Merritt D., Spurzem R., Bischof H.-P., 2006, ApJ, 642, L21

Bessell M. S., 1990, PASP, 102, 1181

Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Blandford R. D., Begelman M. C., 2004, MNRAS, 349, 68

Blandford R. D., Znajek R. L., 1977, MNRAS, 179, 433

Bogdanović T., Eracleous M., Sigurdsson S., 2009, New A Rev., 53, 113

Böhm-Vitense E., 1958, ZAp, 46, 108

Böhm-Vitense E., 1992, Introduction to stellar astrophysics. Volume 3. Stellar structure and evolution.

Bournaud F. et al., 2008, A&A, 486, 741

Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, ApJ, 741, L33

Bournaud F. et al., 2014, ApJ, 780, 57

Boylan-Kolchin M., Ma C.-P., Quataert E., 2005, MNRAS, 362, 184

Brent R., 1973, Algorithms for Minimization Without Derivatives, Dover Books on Mathematics. Dover Publications

Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127

Brinchmann J. et al., 1998, ApJ, 499, 112

Bromm V., Coppi P. S., Larson R. B., 1999, ApJ, 527, L5

Bromm V., Ferrara A., Coppi P. S., Larson R. B., 2001, MNRAS, 328, 969

Bromm V., Loeb A., 2003, ApJ, 596, 34

Browning M. K., 2008, ApJ, 676, 1262

Browning M. K., Brun A. S., Toomre J., 2004, ApJ, 601, 512

Bryan G. L., Norman M. L., 1998, ApJ, 495, 80

Buitrago F., Trujillo I., Conselice C. J., Häußler B., 2013, MNRAS, 428, 1460

Bundy K. et al., 2008, ApJ, 681, 931

Callegari S., Mayer L., Kazantzidis S., Colpi M., Governato F., Quinn T., Wadsley J., 2009, ApJ, 696, L89

Cassinelli J. P., Castor J. I., 1973, ApJ, 179, 189

Ceverino D., Dekel A., Bournaud F., 2010, MNRAS, 404, 2151

Ceverino D., Dekel A., Mandelker N., Bournaud F., Burkert A., Genzel R., Primack J., 2012, MNRAS, 420, 3490

Chabrier G., Gallardo J., Baraffe I., 2007, A&A, 472, L17

Chandrasekhar S., 1943, ApJ, 97, 255

Chapon D., Mayer L., Teyssier R., 2013, MNRAS, 429, 3114

Choi J.-H., Shlosman I., Begelman M. C., 2013, ApJ, 774, 149

Choi J.-H., Shlosman I., Begelman M. C., 2015, MNRAS, 450, 4411

Clark P. C., Glover S. C. O., Smith R. J., Greif T. H., Klessen R. S., Bromm V., 2011, Science, 331, 1040

Cohen S. D., Hindmarsh A. C., Dubois P. F., 1996, Computers in Physics, 10, 138

Coles P., Lucchin F., 2002, Cosmology: The Origin and Evolution of Cosmic Structure, Second Edition. p. 512

Colpi M., Dotti M., 2011, Advanced Science Letters, 4, 181

Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720

Comerford J. M., Greene J. E., 2014, ApJ, 789, 112

Comerford J. M., Pooley D., Barrows R. S., Greene J. E., Zakamska N. L., Madejski G. M., Cooper M. C., 2015, ApJ, 806, 219

Comerford J. M., Schluns K., Greene J. E., Cool R. J., 2013, ApJ, 777, 64

Conselice C. J., Blackburne J. A., Papovich C., 2005, ApJ, 620, 564

Coughlin E. R., Begelman M. C., 2014, ApJ, 781, 82

Courteau S., de Jong R. S., Broeils A. H., 1996, ApJ, 457, L73

Courteau S., Widrow L. M., McDonald M., Guhathakurta P., Gilbert K. M., Zhu Y., Beaton R. L., Majewski S. R., 2011, ApJ, 739, 20

Cox T. J., Dutta S. N., Di Matteo T., Hernquist L., Hopkins P. F., Robertson B., Springel V., 2006, ApJ, 650, 791

Croton D. J. et al., 2006, MNRAS, 365, 11

Daddi E. et al., 2010, ApJ, 713, 686

Davé R., 2009, in Astronomical Society of the Pacific Conference Series, Vol. 419, Galaxy Evolution: Emerging Insights and Future Challenges, Jogee S., Marinova I., Hao L., Blanc G. A., eds., p. 347

Davies R. I., Tacconi L. J., Genzel R., 2004, ApJ, 602, 148

De Breuck C. et al., 2014, A&A, 565, A59

de Vaucouleurs G., de Vaucouleurs A., Corwin, Jr. H. G., Buta R. J., Paturel G., Fouqué P., 1991, Third Reference Catalogue of Bright Galaxies. Volume I: Explanations and references. Volume II: Data for galaxies between 0^h and 12^h. Volume III: Data for galaxies between 12^h and 24^h.

Debattista V. P., Mayer L., Carollo C. M., Moore B., Wadsley J., Quinn T., 2006, ApJ, 645, 209

Dekel A., Birnboim Y., 2006, MNRAS, 368, 2

Dekel A., Birnboim Y., 2008, MNRAS, 383, 119

Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785

del Valle L., Escala A., Maureira-Fredes C., Molina J., Cuadra J., Amaro-Seoane P., 2015, ApJ, 811, 59

Devecchi B., Volonteri M., 2009, ApJ, 694, 302

Devecchi B., Volonteri M., Rossi E. M., Colpi M., Portegies Zwart S., 2012, MNRAS, 421, 1465

Di Matteo P., Combes F., Melchior A.-L., Semelin B., 2007, A&A, 468, 61

Di Matteo T., Khandai N., DeGraf C., Feng Y., Croft R. A. C., Lopez J., Springel V., 2012, ApJ, 745, L29

Dijkstra M., Ferrara A., Mesinger A., 2014, MNRAS, 442, 2036

Dijkstra M., Haiman Z., Mesinger A., Wyithe J. S. B., 2008, MNRAS, 391, 1961

D'Onghia E., Vogelsberger M., Hernquist L., 2013, ApJ, 766, 34

Dotan C., Rossi E. M., Shaviv N. J., 2011, MNRAS, 417, 3035

Dotti M., Colpi M., Haardt F., Mayer L., 2007, MNRAS, 379, 956

Dotti M., Merloni A., Montuori C., 2015, MNRAS, 448, 3603

Dotti M., Ruszkowski M., Paredi L., Colpi M., Volonteri M., Haardt F., 2009, MNRAS, 396, 1640

Downes D., Solomon P. M., 1998, ApJ, 507, 615

Dressler A., 1980, ApJ, 236, 351

Driver S. P. et al., 2006, MNRAS, 368, 414

Driver S. P., Windhorst R. A., Ostrander E. J., Keel W. C., Griffiths R. E., Ratnatunga K. U., 1995, ApJ, 449, L23

Dubinski J., Berentzen I., Shlosman I., 2009, ApJ, 697, 293

Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, MNRAS, 433, 3297

Edge D. O., Shakeshaft J. R., McAdam W. B., Baldwin J. E., Archer S., 1959, MmRAS, 68, 37

Eisenstein D. J., Hu W., 1999, ApJ, 511, 5

Eisenstein D. J. et al., 2011, AJ, 142, 72

Elmegreen B. G., Elmegreen D. M., Fernandez M. X., Lemonias J. J., 2009, ApJ, 692, 12

Eracleous M., Boroson T. A., Halpern J. P., Liu J., 2012, ApJS, 201, 23

Escala A., Larson R. B., 2008, ApJ, 685, L31

Escala A., Larson R. B., Coppi P. S., Mardones D., 2004, ApJ, 607, 765

Escala A., Larson R. B., Coppi P. S., Mardones D., 2005, ApJ, 630, 152

Fabbiano G., Wang J., Elvis M., Risaliti G., 2011, Nature, 477, 431

Fabian A. C., 2012, ARA&A, 50, 455

Fall S. M., 1983, in IAU Symposium, Vol. 100, Internal Kinematics and Dynamics of Galaxies, Athanassoula E., ed., pp. 391–398

Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

Fall S. M., Romanowsky A. J., 2013, ApJ, 769, L26

Fan X. et al., 2006, AJ, 131, 1203

Featherstone N. A., Miesch M. S., 2015, ApJ, 804, 67

Feldmann R., Carollo C. M., Mayer L., 2011, ApJ, 736, 88

Feldmann R., Carollo C. M., Mayer L., Renzini A., Lake G., Quinn T., Stinson G. S., Yepes G., 2010, ApJ, 709, 218

Feldmann R., Mayer L., 2014, arXiv:1404.3212

Feldmann R., Mayer L., Carollo C. M., 2008, ApJ, 684, 1062

Ferrara A., Loeb A., 2013, MNRAS, 431, 2826

Ferrara A., Salvadori S., Yue B., Schleicher D., 2014, MNRAS, 443, 2410

Ferrarese L., Ford H., 2005, Space Sci. Rev., 116, 523

Ferrarese L., Merritt D., 2000, ApJ, 539, L9

Ferraro F. R., Valenti E., Straniero O., Origlia L., 2006, ApJ, 642, 225

Fiacconi D., Mapelli M., Ripamonti E., Colpi M., 2012, MNRAS, 425, 2255

Fiacconi D., Mayer L., Roškar R., Colpi M., 2013, ApJ, 777, L14

Fisher D. B., Drory N., 2008, AJ, 136, 773

Fisher D. B., Drory N., 2011, ApJ, 733, L47

Fisher D. B., Drory N., Fabricius M. H., 2009, ApJ, 697, 630

Fowler W. A., 1966, ApJ, 144, 180

Gavazzi G., Fumagalli M., Cucciati O., Boselli A., 2010, A&A, 517, A73

Gavazzi G. et al., 2013, A&A, 553, A90

Genzel R. et al., 2006, Nature, 442, 786

Gill S. P. D., Knebe A., Gibson B. K., 2004, MNRAS, 351, 399

Gingold R. A., Monaghan J. J., 1977, MNRAS, 181, 375

Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, A&AS, 141, 371

Girardi L. et al., 2010, ApJ, 724, 1030

Governato F. et al., 2010, Nature, 463, 203

Governato F. et al., 2004, ApJ, 607, 688

Graham A. W., Worley C. C., 2008, MNRAS, 388, 1708

Greene J. E. et al., 2010, ApJ, 721, 26

Greenstein J. L., 1963, Nature, 197, 1041

Greenstein J. L., Schmidt M., 1964, ApJ, 140, 1

Greif T. H., Springel V., White S. D. M., Glover S. C. O., Clark P. C., Smith R. J., Klessen R. S., Bromm V., 2011, ApJ, 737, 75

Guedes J., Callegari S., Madau P., Mayer L., 2011, ApJ, 742, 76

Guedes J., Mayer L., Carollo M., Madau P., 2013, ApJ, 772, 36

Gültekin K. et al., 2009, ApJ, 698, 198

Gunn J. E., Gott, III J. R., 1972, ApJ, 176, 1

Gutiérrez C. M., Trujillo I., Aguerri J. A. L., Graham A. W., Caon N., 2004, ApJ, 602, 664

Haardt F., Madau P., 1996, ApJ, 461, 20

Hayashi C., 1961, PASJ, 13, 450

Hernquist L., 1990, ApJ, 356, 359

Hernquist L., 1993, ApJS, 86, 389

Hernquist L., Katz N., 1989, ApJS, 70, 419

Hernquist L., Quinn P. J., 1988, ApJ, 331, 682

Hernquist L., Quinn P. J., 1989, ApJ, 342, 1

Hernquist L., Spergel D. N., Heyl J. S., 1993, ApJ, 416, 415

Hilz M., Naab T., Ostriker J. P., 2013, MNRAS, 429, 2924

Hindmarsh A. C., Brown P. N., Grant K. E., Lee S. L., Serban R., Shumaker D. E., Woodward C. S., 2005, ACM Trans. Math. Softw., 31, 363

Hobbs G. et al., 2010, Classical and Quantum Gravity, 27, 084013

Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 2002, ArXiv Astrophysics e-prints

Holley-Bockelmann K., Khan F. M., 2015, ApJ, 810, 139

Holzer T. E., Axford W. I., 1970, ARA&A, 8, 31

Hopkins P. F. et al., 2010, ApJ, 715, 202

Hopkins P. F., Cox T. J., Hernquist L., Narayanan D., Hayward C. C., Murray N., 2013a, MNRAS, 430, 1901

Hopkins P. F., Keres D., Onorbe J., Faucher-Giguere C.-A., Quataert E., Murray N., Bullock J. S., 2013b, arXiv:1311.2073

Hopkins P. F., Kereš D., Murray N., Quataert E., Hernquist L., 2012, MNRAS, 427, 968

Hopkins P. F., Richards G. T., Hernquist L., 2007, ApJ, 654, 731

Hosokawa T., Omukai K., Yorke H. W., 2012, ApJ, 756, 93

Hosokawa T., Yorke H. W., Inayoshi K., Omukai K., Yoshida N., 2013, ApJ, 778, 178

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Hubble E. P., 1926, ApJ, 64, 321

Huntley J. M., Saslaw W. C., 1975, ApJ, 199, 328

Igumenshchev I. V., Abramowicz M. A., 1999, MNRAS, 303, 309

Ilbert O. et al., 2006, A&A, 453, 809

Ilbert O. et al., 2013, A&A, 556, A55

Jahnke K., Macciò A. V., 2011, ApJ, 734, 92

Kato M., 1983, PASJ, 35, 33

Kazantzidis S., Łokas E. L., Mayer L., 2013, ApJ, 764, L29

Kazantzidis S., Magorrian J., Moore B., 2004, ApJ, 601, 37

Kazantzidis S. et al., 2005, ApJ, 623, L67

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Kerr R. P., 1963, Phys. Rev. Lett., 11, 237

Khan F. M., Holley-Bockelmann K., Berczik P., Just A., 2013, ApJ, 773, 100

Khan F. M., Just A., Merritt D., 2011, ApJ, 732, 89

Khochfar S., Burkert A., 2006, A&A, 445, 403

King A., 2003, ApJ, 596, L27

Kitchatinov L. L., Ruediger G., 1995, A&A, 299, 446

Klessen R. S., Glover S. C. O., Clark P. C., Greif T. H., Bromm V., Jappsen A.-K., 2010, in American Institute of Physics Conference Series, Vol. 1294, American Institute of Physics Conference Series, Whalen D. J., Bromm V., Yoshida N., eds., pp. 28–33

Klessen R. S., Spaans M., Jappsen A.-K., 2007, MNRAS, 374, L29

Knollmann S. R., Knebe A., 2009, ApJS, 182, 608

Komossa S., Burwitz V., Hasinger G., Predehl P., Kaastra J. S., Ikebe Y., 2003, ApJ, 582, L15

Kormendy J., Kennicutt, Jr. R. C., 2004, ARA&A, 42, 603

Koss M. et al., 2014, MNRAS, 445, 515

Kraljic K., Bournaud F., Martig M., 2012, ApJ, 757, 60

Kroupa P., 1998, MNRAS, 298, 231

Kroupa P., Tout C. A., Gilmore G., 1993, MNRAS, 262, 545

Latif M. A., Schleicher D. R. G., Schmidt W., Niemeyer J., 2013, MNRAS, 433, 1607

Lee K.-S. et al., 2011, ApJ, 733, 99

Leitner S. N., Kravtsov A. V., 2011, ApJ, 734, 48

Lin D. N. C., Papaloizou J., 1986, ApJ, 309, 846

Liu X., Civano F., Shen Y., Green P., Greene J. E., Strauss M. A., 2013, ApJ, 762, 110

Liu X., Shen Y., Strauss M. A., Hao L., 2011, ApJ, 737, 101

Lodato G., Natarajan P., 2006, MNRAS, 371, 1813

Łokas E. L., Kazantzidis S., Klimentowski J., Mayer L., Callegari S., 2010, ApJ, 708, 1032

Lucy L. B., 1977, AJ, 82, 1013

Madau P., Haardt F., Dotti M., 2014, ApJ, 784, L38

Madau P., Rees M. J., 2001, ApJ, 551, L27

Magorrian J. et al., 1998, AJ, 115, 2285

Makino J., Funato Y., 2004, ApJ, 602, 93

Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765

Marconi A., Hunt L. K., 2003, ApJ, 589, L21

Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, A&A, 482, 883

Mashchenko S., Wadsley J., Couchman H. M. P., 2008, Science, 319, 174

Mastropietro C., Moore B., Mayer L., Debattista V. P., Piffaretti R., Stadel J., 2005, MNRAS, 364, 607

Mayer L., 2012, in Astronomical Society of the Pacific Conference Series, Vol. 453, Advances in Computational Astrophysics: Methods, Tools, and Outcome, Capuzzo-Dolcetta R., Limongi M., Tornambè A., eds., p. 289

Mayer L., 2013, Classical and Quantum Gravity, 30, 244008

Mayer L., Fiacconi D., Bonoli S., Quinn T., Roškar R., Shen S., Wadsley J., 2015, ApJ, 810, 51

Mayer L., Governato F., Colpi M., Moore B., Quinn T., Wadsley J., Stadel J., Lake G., 2001, ApJ, 559, 754

Mayer L., Governato F., Kaufmann T., 2008, Advanced Science Letters, 1, 7

Mayer L., Kazantzidis S., Escala A., 2008, Mem. Soc. Astron. Italiana, 79, 1284

Mayer L., Kazantzidis S., Escala A., Callegari S., 2010, Nature, 466, 1082

Mayer L., Kazantzidis S., Madau P., Colpi M., Quinn T., Wadsley J., 2007, Science, 316, 1874

Mayer M., Duschl W. J., 2005, MNRAS, 358, 614

McCarthy I. G., Schaye J., Font A. S., Theuns T., Frenk C. S., Crain R. A., Dalla Vecchia C., 2012, MNRAS, 427, 379

McConnell N. J., Ma C.-P., 2013, ApJ, 764, 184

McKee C. F., Ostriker J. P., 1977, ApJ, 218, 148

Medling A. M. et al., 2014, ApJ, 784, 70

Meszaros P., 1974, A&A, 37, 225

Miesch M. S., Brun A. S., Toomre J., 2006, ApJ, 641, 618

Milosavljević M., Merritt D., 2001, ApJ, 563, 34

Milosavljević M., Merritt D., 2003, ApJ, 596, 860

Minkowski R., 1960, ApJ, 132, 908

Misner C. W., Thorne K. S., Wheeler J. A., 1973, Gravitation

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution

Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319

Monaghan J. J., 1992, ARA&A, 30, 543

Moody C. E., Guo Y., Mandelker N., Ceverino D., Mozena M., Koo D. C., Dekel A., Primack J., 2014, arXiv:1405.5266

Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nature, 379, 613

Moore B., Lake G., Katz N., 1998, ApJ, 495, 139

Mortlock A. et al., 2013, MNRAS, 433, 1185

Mortlock D. J. et al., 2011, Nature, 474, 616

Moster B. P., Macciò A. V., Somerville R. S., Naab T., Cox T. J., 2011, MNRAS, 415, 3750

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Munshi F. et al., 2013, ApJ, 766, 56

Muzzin A. et al., 2013, ApJ, 777, 18

Naab T., Burkert A., 2003, ApJ, 597, 893

Naab T., Khochfar S., Burkert A., 2006, ApJ, 636, L81

Naab T., Trujillo I., 2006, MNRAS, 369, 625

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Navarro J. F., Steinmetz M., 2000, ApJ, 538, 477

Noguchi M., 1999, ApJ, 514, 77

Oesch P. A. et al., 2010, ApJ, 714, L47

Oka T., Hasegawa T., Sato F., Tsuboi M., Miyazaki A., Sugimoto M., 2001, ApJ, 562, 348

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 406, 2325

Ostriker E. C., 1999, ApJ, 513, 252

Owocki S. P., Gayley K. G., Shaviv N. J., 2004, ApJ, 616, 525

Palmieri R., Piotto G., Saviane I., Girardi L., Castellani V., 2002, A&A, 392, 115

Pérez-González P. G. et al., 2008, ApJ, 675, 234

Peterson B. M., 1993, PASP, 105, 247

Peterson B. M., 2014, Space Sci. Rev., 183, 253

Planck Collaboration et al., 2015, arXiv:1502.01589

Pontzen A., Roškar R., Stinson G., Woods R., 2013, pynbody: N-Body/SPH analysis for python. Astrophysics Source Code Library

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2002, Numerical recipes in C++ : the art of scientific computing

Prialnik D., 2009, An Introduction to the Theory of Stellar Structure and Evolution

Quataert E., Gruzinov A., 2000, ApJ, 539, 809

Quinlan G. D., Shapiro S. L., 1990, ApJ, 356, 483

Quinn T., Paczynski B., 1985, ApJ, 289, 634

Raiteri C. M., Villata M., Navarro J. F., 1996, A&A, 315, 105

Read J. I., Goerdt T., Moore B., Pontzen A. P., Stadel J., Lake G., 2006, MNRAS, 373, 1451

Reed D., Gardner J., Quinn T., Stadel J., Fardal M., Lake G., Governato F., 2003, MNRAS, 346, 565

Reines A. E., Greene J. E., Geha M., 2013, ApJ, 775, 116

Reines A. E., Sivakoff G. R., Johnson K. E., Brogan C. L., 2011, Nature, 470, 66

Reinhold T., Reiners A., Basri G., 2013, A&A, 560, A4

Rezzolla L., 2009, Classical and Quantum Gravity, 26, 094023

Roberts M. S., Haynes M. P., 1994, ARA&A, 32, 115

Rodriguez C., Taylor G. B., Zavala R. T., Peck A. B., Pollack L. K., Romani R. W., 2006, ApJ, 646, 49

Romanowsky A. J., Fall S. M., 2012, ApJS, 203, 17

Salpeter E. E., 1955, ApJ, 121, 161

Salpeter E. E., 1964, ApJ, 140, 796

Sandage A., 1961, The Hubble atlas of galaxies

Sandage A., 1965, ApJ, 141, 1560

Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749

Sani E., Marconi A., Hunt L. K., Risaliti G., 2011, MNRAS, 413, 1479

Savaglio S. et al., 2005, ApJ, 635, 260

Scannapieco C., Gadotti D. A., Jonsson P., White S. D. M., 2010, MNRAS, 407, L41

Scannapieco E., Silk J., Bouwens R., 2005, ApJ, 635, L13

Scarlata C. et al., 2007, ApJS, 172, 406

Schawinski K., Simmons B. D., Urry C. M., Treister E., Glikman E., 2012, MNRAS, 425, L61

Schmidt M., 1963, Nature, 197, 1040

Schneider A., 2015, MNRAS, 451, 3117

Schönberg M., Chandrasekhar S., 1942, ApJ, 96, 161

Schwarzschild K., 1916, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 1916, Seite 189-196

Scodeggio M., Gavazzi G., Franzetti P., Boselli A., Zibetti S., Pierini D., 2002, A&A, 384, 812

Sellwood J. A., 1989, MNRAS, 238, 115

Sesana A., Haardt F., Madau P., 2007, ApJ, 660, 546

Seyfert C. K., 1943, ApJ, 97, 28

Shaviv N. J., 2001a, ApJ, 549, 1093

Shaviv N. J., 2001b, MNRAS, 326, 126

Shen S., Madau P., Aguirre A., Guedes J., Mayer L., Wadsley J., 2012, ApJ, 760, 50

Shen S., Wadsley J., Stinson G., 2010, MNRAS, 407, 1581

Shibata M., Shapiro S. L., 2002, ApJ, 572, L39

Sijacki D., Springel V., Haehnelt M. G., 2011, MNRAS, 414, 3656

Silk J., Rees M. J., 1998, A&A, 331, L1

Smirnova A., Moiseev A., 2010, MNRAS, 401, 307

Smith B. J., Struck C., Hancock M., Appleton P. N., Charmandaris V., Reach W. T., 2007, AJ, 133, 791

Spaans M., Silk J., 2000, ApJ, 538, 115

Spergel D. N. et al., 2007, ApJS, 170, 377

Springel V., 2005, MNRAS, 364, 1105

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

Springel V. et al., 2005, Nature, 435, 629

Stadel J. G., 2001, PhD thesis, UNIVERSITY OF WASHINGTON

Steinmetz M., Navarro J. F., 2002, New A, 7, 155

Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074

Stinson G. S., Brook C., Macciò A. V., Wadsley J., Quinn T. R., Couchman H. M. P., 2013, MNRAS, 428, 129

Stone J. M., Pringle J. E., Begelman M. C., 1999, MNRAS, 310, 1002

Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253

Tacconi L. J. et al., 2010, Nature, 463, 781

Tamburello V., Capelo P. R., Mayer L., Bellovary J. M., Wadsley J., 2016, arXiv:1603.00021

Tanaka T., Haiman Z., 2009, ApJ, 696, 1798

Tasker E. J., Tan J. C., 2009, ApJ, 700, 358

Teyssier R., Moore B., Martizzi D., Dubois Y., Mayer L., 2011, MNRAS, 414, 195

Thacker R. J., Couchman H. M. P., 2001, ApJ, 555, L17

Thompson M. J., Christensen-Dalsgaard J., Miesch M. S., Toomre J., 2003, ARA&A, 41, 599

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709

Tomczak A. R. et al., 2014, ApJ, 783, 85

Toomre A., 1981, in Structure and Evolution of Normal Galaxies, Fall S. M., Lynden-Bell D., eds., pp. 111–136

Toomre A., Toomre J., 1972, ApJ, 178, 623

Treister E., Schawinski K., Volonteri M., Natarajan P., 2013, ApJ, 778, 130

Tremaine S. et al., 2002, ApJ, 574, 740

Urry C. M., Padovani P., 1995, PASP, 107, 803

Valiante R., Schneider R., Volonteri M., Omukai K., 2016, MNRAS, 457, 3356

van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, MNRAS, 414, 2458

van der Wel A., 2008, ApJ, 675, L13

van Dokkum P. G. et al., 2013, ApJ, 771, L35

Vasiliev E., Antonini F., Merritt D., 2015, ApJ, 810, 49

Volonteri M., 2010, A&A Rev., 18, 279

Volonteri M., Begelman M. C., 2010, MNRAS, 409, 1022

Volonteri M., Silk J., Dubus G., 2015, ApJ, 804, 148

Wadsley J. W., Stadel J., Quinn T., 2004, New A, 9, 137

Wadsley J. W., Veeravalli G., Couchman H. M. P., 2008, MNRAS, 387, 427

Warren M. S., Quinn P. J., Salmon J. K., Zurek W. H., 1992, ApJ, 399, 405

Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230

Weidemann V., 1987, A&A, 188, 74

Weigel A. K., Schawinski K., Treister E., Urry C. M., Koss M., Trakhtenbrot B., 2015, MNRAS, 448, 3167

Weinzirl T., Jogee S., Khochfar S., Burkert A., Kormendy J., 2009, ApJ, 696, 411

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Widrow L. M., Perrett K. M., Suyu S. H., 2003, ApJ, 588, 311

Willott C. J. et al., 2010, AJ, 140, 546

Woosley S. E., Heger A., Weaver T. A., 2002, Reviews of Modern Physics, 74, 1015

Wu X.-B. et al., 2015, Nature, 518, 512

Yang X., Mo H. J., van den Bosch F. C., Bonaca A., Li S., Lu Y., Lu Y., Lu Z., 2013, ApJ, 770, 115

Yoshida N., Omukai K., Hernquist L., Abel T., 2006, ApJ, 652, 6

Zel'dovich Y. B., 1964, Soviet Physics Doklady, 9, 195

Zel'dovich Y. B., 1970, A&A, 5, 84

Żytkow A., 1972, Acta Astron., 22, 103

Curriculum Vitæ

Davide FIACCONI
Birthday: 19 July 1988
Nationality: Italian

Work Experience

- **PhD Student & Teaching Assistant:** PhD student at the Institute for Computational Science under the supervision of Prof. Dr. Lucio Mayer. Teaching assistant for several courses at the BoS and MoS level in Physics and Astrophysics.
  University of Zurich, September 2012 – Present.
- **Content reviewer:** Content reviewer of the book Numpy Beginner's Guide by Ivan Idris and edited by Packt Publishing.
  Packt Publishing, January 2015 – June 2015.
- **Scientific referee:** Referee for the journal Monthly Notices of the Royal Astronomical Society.
  MNRAS, April 2015 – Present.

Education

- **High school:** Liceo Scientifico A. Banfi, Vimercate (MB), Italy
  Degree: *Diploma di Maturità Scientifica* (high-school diploma), 90/100
  Year of graduation: 2007
- **University:** Università di Milano-Bicocca, Milan, Italy
  Degree: *Laurea Triennale in Fisica* (Bachelor of Science, Physics), 110/110 cum Laude
  Year of graduation: 2010
  Thesis title: Simulazioni numeriche di galassie ad anello: il caso di AM0644-741 (Numerical simulations of ring galaxies: the case of AM0644-741)
- **University:** Università di Milano-Bicocca, Milan, Italy
  Degree: *Laurea Magistrale in Astrofisica e Fisica dello Spazio* (Master of Science, Astrophysics and Space Physics), 110/110 cum Laude
  Year of graduation: 2012
  Thesis title: The fate of collisional ring galaxies

Conferences, Schools & Colloquia

- **School**, Early Galaxy Formation in LCDM Cosmology, Jerusalem Winter School, 2013

- **School**, Star formation in galaxy evolution, 43^{rd} Saas-Fee Course, 2013
- **Conference**, Massive Black Holes: Birth, Growth and Impact, KITP, 2013
  Talk: Evolution of massive black hole pairs in clumpy environments: from circumnuclear disks to major mergers
- **Conference**, Exascale Computing in Astrophysics, Centro Stefano Franscini, 2013
- **Conference**, Nuclear clusters in galaxies, and the role of environment, Lorentz Center, 2014
  Talk: Formation of nuclear star cluster in high-z galaxy mergers: implications for massive black hole formation
- **Conference**, Good Sense and Dominant Ideology in Galaxy & Planet Formation, Centro Stefano Franscini, 2014
- **Colloquium**, eLISA Consortium Meeting, University of Milan-Bicocca, 2014
  Talk: Gas-rich major mergers: supermassive black hole formation and dynamics
- **Conference**, Unveiling the AGN-Galaxy evolution connection, Puerto Varas, 2015
  Talk: Formation of supermassive black holes in high-redshift galaxy major merger
- **Conference**, Star formation history of the Universe, MIAPP, 2015
  Talk: The development of galaxy morphology: a high-redshift perspective
- **Colloquium**, Astrophysics Colloquium, Capodimonte Astronomical Observatory, 2015
  Talk: The formation of massive black hole seeds in high-redshift galaxy mergers
- **Colloquium**, YAGN15 Meeting, University of Milan-Bicocca, 2015
  Talk: Forming massive black hole seeds via quasi-stars

Publications

- Khan, F., **Fiacconi, D.**, Mayer, L., Berczik, P., & Just, A., *Swift coalescence of supermassive black holes in cosmological mergers of massive galaxies*, submitted to ApJ Letters, arXiv:1604.00015
- **Fiacconi, D.**, & Rossi, E. M., *Light or heavy supermassive black hole seeds: the role of internal rotation in the fate of supermassive stars*, submitted to MNRAS, arXiv:1604.03936
- **Fiacconi, D.**, Madau, P., Potter, D., & Stadel, J., *Cold dark matter substructures in elliptical galaxy halos*, ApJ, in press, arXiv:1602.03526
- Lupi, A., Haardt, F., Dotti, M., **Fiacconi, D.**, Mayer, L., & Madau, P., *Growing massive black holes through super-critical accretion of stellar-mass seeds*, MNRAS, 456, 2993
- **Fiacconi, D.**, & Rossi, E. M., *Bright vigorous winds as signposts of supermassive black hole birth*, 2016, MNRAS, 455, 2
- Fanali, R., Dotti, M., **Fiacconi, D.**, & Haardt, F., *Bar-formation as driver of gas inflows in isolated disc galaxies*, 2015, MNRAS, 454, 3641
- Mayer, L., **Fiacconi, D.**, et al., *Direct formation of supermassive black holes in metal-enriched gas at the heart of high-redshift galaxy mergers*, 2015, ApJ, 810, 51
- Roškar, R., **Fiacconi, D.**, et al., *Orbital Decay of Supermassive Black Hole Binaries in Inhomogeneous Multiphase Merger Remnants*, 2015, MNRAS, 449, 494
- **Fiacconi, D.**, Feldmann, R., & Mayer, L., *The Argo simulation – II. The early build-up of the Hubble sequence*, 2015, MNRAS, 446, 1957

- **Fiacconi, D.**, Mayer, L., Roškar, R., & Colpi, M., *Massive Black Hole Pairs in Clumpy, Self-gravitating Circumnuclear Disks: Stochastic Orbital Decay*, 2013, ApJ, 777, L14
- **Fiacconi, D.**, Mapelli, M., Ripamonti, E., & Colpi, M., *Adaptive mesh refinement simulations of collisional ring galaxies: effects of the interaction geometry*, 2012, MNRAS, 425, 2255

Acknowledgements

I would like to express my sincere gratitude to my advisor Professor Dr. Lucio Mayer. I really would like to thank you for encouraging me in all respects, not only in my research but also during the difficult and doubtful moments of my PhD, being always ready to support my choices and to provide an optimistic and certain opinion about every possible pathway ahead of me. Thank you very much also for having never doubted my abilities, even when I was ready to do that!

I have been very lucky during these years to count on other two persons that became scientific models for me: Professor Dr. Elena Maria Rossi, and Professor Dr. Piero Madau. I really enjoyed working with both of you during these years, and I trust that we will continue to produce great science together. Thank you very much for your mentorship and for your continuos support.

I would also like to thank my external reviewers, Professor Dr. Arif Babul, and Professor Dr. Priyamvada Natarajan, for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

Special thanks go to the other PhD students and postdocs who shared these years with me: we had a lot of fun together and I found many good friends among you all. I want to explicitly express my gratitude (in random order) to Pedro, for his continuous support since the first time we met in Santa Barbara; to Simone & Lorenzo, for the (never enough) good time we spent together; to Paweł, for being a good friend, for the nice experiences together at conferences and for the many that will come; to my officemates (Aleksandra, Maria, and Valentina), for being patient with your noisy officemate always swearing against the screen, for our laughs around our four desks, and for having really made me feel “beato tra le donne”; to Irshad, for the nice time we spent together and for the beautiful reunion we had in India together with the whole “Zurich-Delhi delegation” after his PhD defense; to Mihai, for the long and good chats we had together and for the mutual support when the PhD defense was approaching; to Davide & Volker, who first “received” me as a new PhD student when I arrived at the institute. I want to acknowledge the invaluable help of Esther, Regina, and Suzanne during these years. I also want to thank my flatmate Francesco and his mother Concetta for being a good friend and for having treated me as part of the family.

Un Grazie speciale è riservato alla mia famiglia: mia Mamma, mio Papà e mia sorella Cecilia. Nonostante la distanza geografica, avete sempre rappresentato un supporto incondizionato e un'inesauribile fonte di consiglio per combattere le difficoltà che hanno afflitto il mio animo durante questi anni, e allo stesso tempo siete stati sempre pronti ad incitarmi verso il raggiungimento di qualunque successo e a condividerne con me le gioie.

Vorrei concludere ringraziando la persona che forse più di tutte ha vissuto assieme a me le gioie ed i dolori di questo periodo, la mia Anna Chiara. Questa esperienza non si sarebbe conclusa tanto bene se non fosse stato per la tua infinita pazienza nel sopportare la distanza e i miei malumori, essendo al contempo in grado di supportarmi continuamente e di infondermi gioia e fiducia nelle mie capacità, anteponendo sempre il tentativo di regalarmi un sorriso nei miei momenti più cupi. Grazie di esserci sempre per me, Grazie di rappresentare l'ultimo scoglio al quale aggrapparsi in un mare in tempesta, Grazie di essere il motore del mio entusiasmo per la vita e il motivo per provare ad essere una persona sempre migliore.

Grazie.